

Process convergence for the complexity of Radix Selection on Markov sources

Kevin Leckey
School of Mathematical Sciences
Monash University, VIC 3800
Melbourne
Australia

Ralph Neininger*
Institute for Mathematics
Goethe University Frankfurt
60054 Frankfurt am Main
Germany

Henning Sulzbach†
School of Computer Science
McGill University
H3A 2K6, Montreal
Canada

May 10, 2016

Abstract

A fundamental algorithm for selecting ranks from a finite subset of an ordered set is Radix Selection. This algorithm requires the data to be given as strings of symbols over an ordered alphabet, e.g., binary expansions of real numbers. Its complexity is measured by the number of symbols that have to be read. In this paper the model of independent data identically generated from a Markov chain is considered.

The complexity is studied as a stochastic process indexed by the ranks to be selected. The orders of mean and variance of the complexity and limit theorems are derived. For uniform data and the asymmetric Bernoulli model, we find weak convergence of the appropriately normalized complexity towards a Gaussian process with explicit mean and covariance functions in the space of càdlàg functions with the Skorokhod metric. For all other Markov sources we show that such a convergence does not hold. We also study two further models for the ranks: uniformly chosen ranks and the worst case rank complexities which are of interest in computer science.

AMS 2010 subject classifications. Primary 60F17, 60G15 secondary 68P10, 60C05, 68Q25.

Key words. Radix Selection, Gaussian process, Markov source model, complexity, weak convergence, probabilistic analysis of algorithms.

1 Introduction

In the probabilistic analysis of algorithms the complexity of fundamental algorithms is studied under models of random input. This allows to describe the typical behavior of an algorithm and is often more meaningful than the worst case complexity classically considered in computer science. In this paper we study the algorithm Radix Selection on independent strings generated by a Markov source.

*Supported by DFG grant NE 828/2-1

†Supported by the FSMP, reference: ANR-10-LABX-0098, and a Feodor Lynen Research Fellowship of the Alexander von Humboldt Foundation

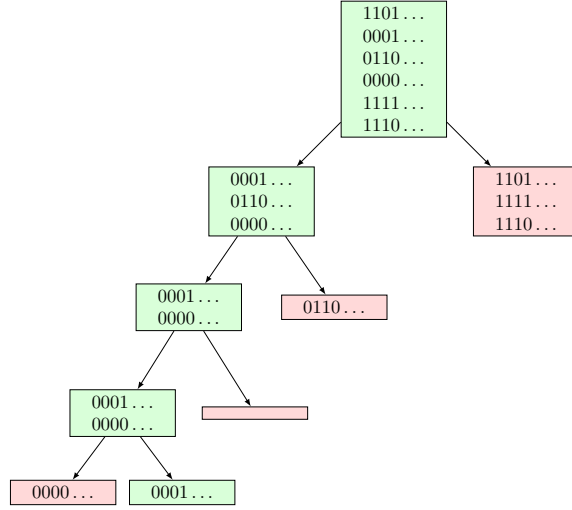


Figure 1: A schematic representation of Radix Selection with $b = 2$ buckets searching for the element of rank 2 in a list of 6 elements given in their binary expansions. Arrows indicate the splitting into buckets, the green color indicates buckets containing the element of rank 2. The total number of bucket operations is $6 + 3 + 2 + 2 = 13$.

Radix Selection selects an order statistic from a set of data in $[0, 1]$ as follows. First, an integer $b \geq 2$ is fixed and the unit interval is decomposed into the intervals, also called *buckets*, $[0, 1/b)$, $[1/b, 2/b)$, \dots , $[(b-2)/b, (b-1)/b)$ and $[(b-1)/b, 1]$. The data are assigned to these buckets according to their values. Note that this just corresponds to group the data according to the first symbols of their b -ary expansions. If the bucket containing the datum with rank to be selected contains further data, the algorithm is recursively applied by decomposing this bucket equidistantly using the integer b . The algorithm stops once the bucket containing the sought rank contains no other data. Assigning a datum to a bucket is called a *bucket operation*, and the algorithm's complexity is measured by the total number of bucket operations required. An illustration of this procedure is given in Figure 1.

Radix Selection is especially suitable when data are stored as expansions in base (radix) b , the case $b = 2$ being the most common on the level of machine data. For such expansions a bucket operation breaks down to access a digit (or bit).

In this paper we study the complexity of Radix Selection in the probabilistic setting that n data are modeled independently with b -ary expansions generated from a Markov chain on the alphabet $\{0, \dots, b-1\}$. For the ranks to be selected we use three models. First, all possible ranks are considered simultaneously. Hence, we study the stochastic process of the complexities indexed by the ranks $1, \dots, n$. We choose a scaling in time and space which asymptotically gives access to the complexity to select quantiles from the data, i.e., ranks of the size tn with $t \in [0, 1]$. We call this model for the ranks the *quantile-model*. Second, we consider the complexity of a random rank uniformly distributed over $\{1, \dots, n\}$ and independent from the data. This is the model proposed and studied (for independent, uniformly over $[0, 1]$ distributed data) in Mahmoud, Flajolet, Jacquet and Régnier [17]. The complexities of all ranks are averaged in this model and, in accordance with the literature, we call it the model of *grand averages*. Third, we study the worst rank complexity. Here, the data are still random and the worst case is taken over the possible ranks $\{1, \dots, n\}$. We call this *worst case rank*.

The main results of this work concern the asymptotic orders of mean and variance as well as limit laws for the complexity of Radix Selection for our Markov source model for all three models of

ranks. The organisation of the paper is roughly as follows: For the *quantile-model* in Section 2, we find Gaussian limit processes for the uniform model (defined below) for the data as well as for the asymmetric Bernoulli model (defined below). Throughout the present paper, convergence towards a Gaussian process takes place in the space $D[0, 1]$ of càdlàg functions endowed with the Skorokhod topology. For the general Markov source model with $b = 2$, we identify the first asymptotic term of the mean complexity and show that process convergence in $D[0, 1]$ does not hold. We discuss *grand averages* in Section 3. For uniform data it was shown in Mahmoud et al. [17] that the normalized complexity is asymptotically normal. We find that for Markov sources (with $b = 2$) other than uniform the limit distribution is no longer normal, and the complexity is less concentrated. Finally, Section 4 covers the *worst case rank* model. For the uniform model and the asymmetric Bernoulli model we find limit laws for the worst case rank complexity. For general Markov sources the linear growth order of the complexity is identified.

A general reference on bucket algorithms is Devroye [5]. A large body of probabilistic analysis of digital structures is based on methods from analytic combinatorics, see Flajolet and Sedgewick [8], Knuth [14] and Szpankowski [19]. For an approach based on renewal theory see Janson [12] and the references given therein. Our Markov source model is a special case of the model of dynamical sources, see Clément, Flajolet and Vallée [4] as well as [10, 3]. A related important model is the density model studied in Devroye [6].

We close this introduction defining the Markov source model explicitly, fixing some standard notation and stating corresponding results for the related Radix Sorting algorithm.

The Markov source model: We model data strings over the alphabet $\Sigma = \{0, \dots, b-1\}$ with a fixed integer $b \geq 2$ generated by a homogeneous Markov chain. A data string $s = (s_i)_{i \geq 1}$ is also interpreted as b -ary expansion of a real number $s \in [0, 1]$ via the identification

$$s = \sum_{i=1}^{\infty} s_i b^{-i}. \quad (1)$$

Conversely, if to $s \in [0, 1]$ a b -ary expansion $s = (s_i)_{i \geq 1}$ is associated, to avoid ambiguity, we require the expansion to satisfy $s_i < b-1$ for infinitely many $i \in \mathbb{N}$. (For $s = 1$, we use the expansion where $s_i = b-1$ for all $i \in \mathbb{N}$.) The most important case is $b = 2$ where the data are binary strings.

In general, a homogeneous Markov chain on Σ is given by its initial distribution $\mu = \sum_{\ell=0}^{b-1} \mu_\ell \delta_\ell$ on Σ and the transition matrix $(p_{ij})_{i,j \in \Sigma}$. Here, δ_x denotes the Dirac measure in $x \in \mathbb{R}$. Hence, the initial state is ℓ with probability μ_ℓ for $\ell = 0, \dots, b-1$. We have $\mu_\ell \in [0, 1]$ and $\sum_{\ell=0}^{b-1} \mu_\ell = 1$. A transition from state i to j happens with probability p_{ij} , $i, j \in \Sigma$. Now, a data string is generated as the sequence of states taken by the Markov chain. In our Markov source model assumed subsequently, all data strings are independent and identically distributed according to the given Markov chain.

We always assume that $p_{ij} < 1$ for all $i, j \in \Sigma$. Note that we do not require the Markov chain to converge to a stationary distribution nor that it starts in a stationary distribution.

The case $p_{ij} = \mu_i = 1/b$ for all $i, j \in \Sigma$ is the case where all symbols within all data are independent and uniformly distributed over Σ . Then the corresponding numbers on $[0, 1]$ associated as in (1) are independent and uniformly distributed over $[0, 1]$. We call this the *uniform model*. For $b = 2$, the uniform model is also called *symmetric Bernoulli model*. The *asymmetric Bernoulli model* for $b = 2$ is the case in which $p_{i1} = \mu_1 = p$ for $i = 0, 1$ and $p \in (0, 1)$ with $p \neq \frac{1}{2}$.

Some results of the present paper have been announced in the extended abstract [15].

Notation. We write \xrightarrow{d} for convergence in distribution and $\stackrel{d}{=}$ for equality in distribution. By $B(n, p)$ with $n \in \mathbb{N}$ and $p \in [0, 1]$ the binomial distribution is denoted, by $B(p)$ the Bernoulli distribution with success probability p , by $\mathcal{N}(\mu, \sigma^2)$ the normal distribution with mean $\mu \in \mathbb{R}$ and variance $\sigma^2 > 0$. Furthermore, $\|X\|_s := (\mathbb{E}[|X|^s])^{1/s}$ denotes the s -norm, $s \in [1, \infty)$, of a real-valued random variable X . We also use the abbreviations $x \wedge y = \min\{x, y\}$ and $x \vee y = \max\{x, y\}$ for real numbers x, y . Finally, the Bachmann–Landau symbols are used.

Radix Sorting. The Radix Sorting algorithm starts by assigning all data to the buckets as for Radix Selection. Then it recurses on all buckets containing more than one datum. Clearly, this leads to a sorting algorithm. The complexity of Radix Sorting is measured by the number of bucket operations. It has been analyzed thoroughly in the uniform model with precise expansions for mean and variance involving periodic functions and a central limit law for the normalized complexity, see Knuth [14], Jacquet and Régnier [11], Kirschenhofer, Prodinger and Szpankowski [13] and Mahmoud et al. [17].

For the Markov source model (with $b = 2$ and $0 < p_{ij} < 1$ for $i, j = 1, 2$) the orders of mean and variance and a central limit theorem for the complexity of Radix Sorting were derived in Leckey, Neininger and Szpankowski [16].

2 The quantile-model

The quantile model is the most refined model for the ranks and the most difficult model to analyze. The results and proofs are organized as follows: Section 2.1 contains the uniform model with general $b \geq 2$ where we obtain a Gaussian limit process in Theorem 2.1. Another Gaussian limit law holds in the asymmetric Bernoulli model (with $b = 2$) which is stated later in Theorem 2.5. The discussion of all other Markov sources is contained in Section 2.2. Here, we find the first order growth of the process and that limit laws in the spirit of Theorems 2.1 and 2.5 do not hold for these Markov sources.

2.1 The uniform model

In this section the data is sampled independently with the uniform distribution on $[0, 1]$. We fix $b \geq 2$ and consider Radix Selection using b buckets in each step. The number $Y_n(\ell)$ of bucket operations needed to select rank $\ell \in \{1, \dots, n\}$ in a set of n data is studied as a process in $1 \leq \ell \leq n$. We write $Y_n := (Y_n(\ell))_{1 \leq \ell \leq n}$. For a refined asymptotic analysis we normalize the process in space and time and consider $X_n = (X_n(t))_{0 \leq t \leq 1}$ defined for $n \geq 1$ and $t \in [0, 1]$ by

$$X_n(t) := \frac{Y_n(\lfloor tn \rfloor + 1) - \frac{b}{b-1}n}{\sqrt{n}}. \quad (2)$$

Here, we set $Y_n(n+1) := Y_n(n)$. The process X_n has càdlàg paths and is considered as a random variable in $D[0, 1]$ endowed with the Skorokhod metric d_{sk} , see Billingsley [1, Chapter 3].

Subsequently, we use prefixes of b -ary expansions. For $s, t \in [0, 1]$ based on their b -ary expansions $s = \sum_{i=1}^{\infty} s_i \cdot b^{-i}$, $t = \sum_{i=1}^{\infty} t_i \cdot b^{-i}$ with $s_i, t_i \in \{0, \dots, b-1\}$ with the conventions stated in the introduction, we denote the length of the longest common prefix by

$$j(s, t) := \max\{i \in \mathbb{N} \mid (s_1, \dots, s_i) = (t_1, \dots, t_i)\}. \quad (3)$$

Here, and subsequently, we apply the conventions $\max \emptyset := 0$ and $\max \mathbb{N} := \infty$.

We have the following functional limit theorem:

Theorem 2.1. *Let $b \in \mathbb{N}$ with $b \geq 2$. Consider Radix Selection using b buckets on a set of independent data uniformly distributed on $[0, 1]$. For the process $X_n = (X_n(t))_{0 \leq t \leq 1}$ of the normalized number of bucket operations $Y_n(\ell)$ as defined in (2) we have weak convergence, as $n \rightarrow \infty$, in $(D[0, 1], d_{sk})$:*

$$X_n \xrightarrow{d} G.$$

Here, $G = (G(t))_{t \in [0, 1]}$ is a centered Gaussian process with covariance function

$$\mathbb{E}[G(s)G(t)] = \frac{b}{(b-1)^2} - \frac{b+1}{(b-1)^2} b^{-j(s, t)}, \quad s, t \in [0, 1], \quad (4)$$

where $j(s, t)$ is the length of the longest common prefix defined in (3) and $b^{-\infty} := 0$.

An alternative characterization of the limit process G is given in Proposition 2.2 below. We first give a rough outline of our proof of Theorem 2.1.

Outline of the analysis: To set up a recurrence for the process $Y_n := (Y_n(\ell))_{1 \leq \ell \leq n}$ we denote by $I^n = (I_1^n, \dots, I_b^n)$ the numbers of elements in the b buckets after distribution of all n elements in the first partitioning stage. We abbreviate $F_0^n := 0$ and

$$F_r^n := \sum_{j=1}^r I_j^n, \quad 1 \leq r \leq b.$$

Note that we have $F_b^n = n$. After the first partitioning phase, the element of rank ℓ is in bucket r if and only if $F_{r-1}^n < \ell \leq F_r^n$. This implies the recurrence

$$Y_n \stackrel{d}{=} \left(\sum_{r=1}^b \mathbf{1}_{(F_{r-1}^n, F_r^n]}(\ell) Y_{I_r^n}^r (\ell - F_{r-1}^n) + n \right)_{1 \leq \ell \leq n}, \quad (5)$$

where $(Y_j^1), \dots, (Y_j^b), I^n$ are independent and the Y_j^r have the same distribution as Y_j for all $j \geq 0$ and $r = 1, \dots, b$.

From the underlying probabilistic model it follows that the vector I^n has the multinomial $M(n; \frac{1}{b}, \dots, \frac{1}{b})$ distribution. Hence, we have $\frac{1}{n} I^n \rightarrow (\frac{1}{b}, \dots, \frac{1}{b})$ almost surely as $n \rightarrow \infty$ and

$$\frac{I^n - \frac{1}{b}(n, \dots, n)}{\sqrt{n}} \xrightarrow{d} (N_1, \dots, N_b),$$

where (N_1, \dots, N_b) is a multivariate normal distribution $\mathcal{N}(0, \Omega)$ with mean zero and covariance matrix Ω given by $\Omega_{ij} = \frac{b-1}{b^2}$ if $i = j$ and $\Omega_{ij} = -\frac{1}{b^2}$ if $i \neq j$. Note that $\sum_{r=1}^b N_r = 0$ almost surely. Below, we denote by

$$\mathcal{N} = (\mathcal{N}_1, \dots, \mathcal{N}_b) \quad (6)$$

a vector with distribution $\frac{b}{b-1}(N_1, \dots, N_b)$. Hence $(\mathcal{N}_1, \dots, \mathcal{N}_b)$ has a multivariate normal distribution with mean zero and covariance matrix $\Upsilon = (\Upsilon_{ij})_{i,j \in \Sigma}$ given by

$$\Upsilon_{ij} = \begin{cases} \frac{1}{b-1}, & \text{if } i = j, \\ -\frac{1}{(b-1)^2}, & \text{if } i \neq j. \end{cases} \quad (7)$$

For the normalized processes X_n in (2) we thus obtain

$$X_n \stackrel{d}{=} \left(\sum_{r=1}^b \mathbf{1}_{\left[\frac{F_{r-1}^n}{n}, \frac{F_r^n}{n}\right)}(t) \sqrt{\frac{I_r^n}{n}} X_{I_r^n}^r \left(\frac{nt - F_{r-1}^n}{I_r^n} \right) + \sum_{r=1}^b \mathbf{1}_{\left[\frac{F_{r-1}^n}{n}, \frac{F_r^n}{n}\right)}(t) \frac{b}{b-1} \frac{I_r^n - \frac{1}{b}n}{\sqrt{n}} \right)_{0 \leq t \leq 1}, \quad (8)$$

with conditions on independence and identical distributions as in (5).

To associate to recurrence (8) a limit equation in the spirit of the contraction method, we introduce a family of parameter transformations: for $0 \leq a < b \leq 1$, let

$$\mathfrak{A}_{a,b} : D[0, 1] \rightarrow D[0, 1], \quad \mathfrak{A}_{a,b}(f)(t) = \mathbf{1}_{[a,b)} f\left(\frac{t-a}{b-a}\right).$$

Here, and subsequently, to simplify notation, we set

$$[a, 1] := [a, 1] \quad \text{for } 0 < a < 1. \quad (9)$$

Moreover, we define $\mathfrak{B} : \mathbb{R}^b \rightarrow D[0, 1]$ with (for $v = (v_1, \dots, v_b)$) by (note convention (9))

$$v \mapsto \mathfrak{B}(v), \quad \mathfrak{B}(v)(t) = \sum_{r=1}^b \mathbf{1}_{\left[\frac{r-1}{b}, \frac{r}{b}\right)}(t) v_r \text{ for } t \in [0, 1].$$

Then we associate the limit equation (again with convention (9))

$$X \stackrel{d}{=} \sum_{r=1}^b \frac{1}{\sqrt{b}} \mathfrak{A}_{\frac{r-1}{b}, \frac{r}{b}}(X^r) + \mathfrak{B}(\mathcal{N}), \quad (10)$$

where $X^1, \dots, X^b, \mathcal{N}$ are independent and X^1, \dots, X^b are distributed like X . Further, \mathcal{N} has the centered multivariate normal distribution with covariance matrix Υ given in (7).

Our proof of the convergence in Theorem 2.1 below relies on a contractive argument for the distance between X_n and an accompanying sequence Q_n connecting X_n to G . In the context of the contraction method, one uses the same idea to infer that G is the unique solution (in distribution) to (10) with càdlàg paths under the condition $\mathbb{E}[\|G\|^3] < \infty$. In fact, we have the following stronger result.

Proposition 2.2. *The process G in Theorem 2.1 is the unique càdlàg process (in distribution) satisfying (10).*

The remaining part of this section is devoted to the proofs of Theorem 2.1 and Proposition 2.2.

Proof of Theorem 2.1. The proof follows the approach developed in [18] leading to Theorem 1.1 there. The starting point is a family of independent and identically distributed random variables $\{(I^{n,\vartheta})_{n \geq 0}, \mathcal{N}^\vartheta : \vartheta \in T\}$ where $T = \bigcup_{n \geq 0} \{1, 2, \dots, b\}^n$ denotes the complete b -ary tree, $I^{n,\vartheta}$ has distribution $M(n; \frac{1}{b}, \dots, \frac{1}{b})$ and \mathcal{N}^ϑ has the distribution given in (6) with

$$\frac{I^{n,\vartheta} - \frac{1}{b}(n, \dots, n)}{\sqrt{n}} \rightarrow \frac{b-1}{b} \mathcal{N}^\vartheta, \quad n \rightarrow \infty, \quad (11)$$

almost surely for all $\vartheta \in T$.

The limit process. Setting $G_0^\vartheta := 0$ for all $\vartheta \in T$, we recursively construct random variables $G_n^\vartheta, n \geq 1, \vartheta \in T$, by (with convention (9))

$$G_{n+1}^\vartheta = \sum_{r=1}^b \frac{1}{\sqrt{b}} \mathfrak{A}_{\frac{r-1}{b}, \frac{r}{b}}(G_n^{\vartheta r}) + \mathfrak{B}(\mathcal{N}^\vartheta).$$

It follows that, for any $p \in \mathbb{N}$,

$$\begin{aligned} \mathbb{E}[\|G_{n+1}^\vartheta - G_n^\vartheta\|^p] &\leq \sum_{r=1}^b b^{-p/2} \mathbb{E}[\|G_n^{\vartheta r} - G_{n-1}^{\vartheta r}\|^p] \\ &= b^{-p/2+1} \mathbb{E}[\|G_n^\vartheta - G_{n-1}^\vartheta\|^p] \leq b^{-n(p/2-1)} \mathbb{E}[\|G_1^\vartheta\|^p]. \end{aligned} \quad (12)$$

From here, choosing $p > 2$, standard arguments (compare e.g. the proof of Lemma 2.1 in [18]) show that, almost surely, G_n^ϑ is uniformly Cauchy. By the completeness of the càdlàg space endowed with the supremum norm, G_n^ϑ is uniformly convergent. The limit denoted by G^ϑ satisfies

$$G^\vartheta = \sum_{r=1}^b \frac{1}{\sqrt{b}} \mathfrak{A}_{\frac{r-1}{b}, \frac{r}{b}}(G^{\vartheta r}) + \mathfrak{B}(\mathcal{N}^\vartheta).$$

In particular, G^ϑ satisfies (10) since $G^\vartheta, \vartheta \in T$ is a family of identically distributed random variables and $G^{\vartheta^1}, \dots, G^{\vartheta^b}$ are independent. From (12), it follows that $\mathbb{E}[\|G_n^\vartheta - G^\vartheta\|^p] \rightarrow 0$ for all $p \in \mathbb{N}$. Inductively, one can easily show that G_n^ϑ is a Gaussian process with zero mean for all $n \geq 1$. Hence, the same follows for G^ϑ . Denoting $\sigma(s, t) = \mathbb{E}[G^\vartheta(t)G^\vartheta(s)]$, using the b -ary expansions $t = t_1 t_2 \dots, s = s_1 s_2 \dots$, and (10), we find

$$\sigma(s, t) = \begin{cases} b^{-1} \sigma(bt - t_1, bs - t_1) + (b-1)^{-1}, & \text{if } j(s, t) \geq 1, \\ -(b-1)^{-2}, & \text{if } j(s, t) = 0. \end{cases}$$

This yields the expression for the covariance function in (4) for $s \neq t$. The variance then follows by right continuity of G^ϑ .

Construction of the discrete process. We continue with the construction of suitable versions of X_n based on the distributional recurrence (8). Let $X_0^\vartheta := X_1^\vartheta := 0$ for all $\vartheta \in T$ and recursively, for $n \geq 2$ and with convention (9),

$$X_n^\vartheta = \sum_{r=1}^b \sqrt{\frac{I_r^{n,\vartheta}}{n}} \mathfrak{A}_{\frac{F_r^{n,\vartheta}}{n-1}, \frac{F_r^{n,\vartheta}}{n}} \left(X_{I_r^{n,\vartheta}}^{\vartheta r} \right) + \sum_{r=1}^b \mathbf{1}_{\left[\frac{F_r^{n,\vartheta}}{n-1}, \frac{F_r^{n,\vartheta}}{n} \right)} \frac{b}{b-1} \frac{I_r^{n,\vartheta} - \frac{1}{b}n}{\sqrt{n}}, \quad (13)$$

Here, $F_r^{n,\vartheta} = \sum_{j=1}^r I_j^{n,\vartheta}$, $1 \leq r \leq b$. Note that this definition is more subtle than the corresponding one in [18], since, with positive probability, we have $I_r^{n,\vartheta} = n$ for some $r = 1, \dots, b$. To resolve this issue, let

$$M = \bigcap_{n \geq 2} \bigcap_{\vartheta \in T} \bigcup_{k \geq 1} \bigcap_{1 \leq r_1, \dots, r_k \leq b} \bigcup_{1 \leq i \leq k} \{I_{r_i}^{n, \vartheta r_1 \dots r_{i-1}} < n\}, \quad (14)$$

and note that $\mathbb{P}(M) = 1$. On M , the definition (13) only involves a finite (but random) number of recursions on the right hand side, thus it defines the family of càdlàg functions $(X_n^\vartheta), n \geq 2, \vartheta \in T$. For convenience, we set $X_n^\vartheta := 0$ for all $n \geq 2, \vartheta \in T$ on the event M^c .

Next, again by induction on n , we prove that X_n^ϑ is measurable for each $n \geq 2, \vartheta \in T$. Assume it was correct for all $m < n$ and $\vartheta \in T$. Then, for a measurable set $B \subseteq D[0, 1]$ with $0 \notin B$, recalling the set M defined in (14), we have

$$\begin{aligned} \{X_n^\vartheta \in B\} &= M \cap \left(\bigcup_{\substack{0 \leq i_1, \dots, i_r \leq n-1 \\ i_1 + \dots + i_r = n}} \{I^{n,\vartheta} = (i_1, \dots, i_r)\} \cap \left\{ \sum_{r=1}^K \sqrt{\frac{i_r}{n}} \mathfrak{A}_{\frac{f_{r-1}}{n-1}, \frac{f_r}{n}} (X_{i_r}^{\vartheta r}) \right. \right. \\ &\quad \left. \left. + \mathbf{1}_{\left[\frac{f_{r-1}}{n-1}, \frac{f_r}{n} \right)} \frac{b}{b-1} \frac{i_r - \frac{1}{b}n}{\sqrt{n}} \in B \right\} \bigcup_{r=1}^K \{I_r^{n,\vartheta} = n\} \cap \{X_n^{\vartheta r} + \sqrt{n} \in B\} \right). \end{aligned}$$

Iterating the argument and using the induction hypothesis, there exist a sequence of increasing measurable sets $(A_k)_{k \geq 1}$ and a sequence of decreasing sets $(B_k)_{k \geq 1}$, such that, for any $k \geq 1$, we have

$$\{X_n^\vartheta \in B\} = M \cap (A_k \cup B_k),$$

where $B_k \subseteq \bigcup_{1 \leq r_1, \dots, r_k \leq b} \bigcap_{i=1}^k \{I_{r_i}^{n, \vartheta r_1 \dots r_{i-1}} = n\}$. Thus,

$$\{X_n^\vartheta \in B\} = M \cap \left(\bigcup_{k \geq 0} A_k \cup \bigcap_{k \geq 1} B_k \right) = M \cap \bigcup_{k \geq 0} A_k$$

since $\bigcap_{k \geq 1} B_k \subseteq M^c$. Using the same ideas for the case $B = \{0\}$, it follows that X_n^ϑ is measurable.

To show that X_n^ϑ has the desired distribution, note that, for the sequence of random variables X_n satisfying (8), we have

$$\begin{aligned} \mathbb{P}(X_n \in B) &= \sum_{\substack{0 \leq i_1, \dots, i_r \leq n-1 \\ i_1 + \dots + i_r = n}} \mathbb{P}(I^{(n)} = (i_1, \dots, i_r)) \mathbb{P}\left(\sum_{r=1}^K \sqrt{\frac{i_r}{n}} \mathfrak{A}_{\frac{f_{r-1}}{n}, \frac{f_r}{n}}(X_{i_r}^r)\right. \\ &\quad \left. + \mathbf{1}_{[\frac{f_{r-1}}{n}, \frac{f_r}{n})} \frac{b}{b-1} \frac{i_r - \frac{1}{b}n}{\sqrt{n}} \in B\right) + \sum_{r=1}^K \mathbb{P}(I_r^{(n)} = n) \mathbb{P}(X_n^r + \sqrt{n} \in B). \end{aligned}$$

Using the analogous formula for the process X_n^ϑ and assuming that $X_m^\vartheta \stackrel{d}{=} X_m$ for $m < n$, we have

$$\sup_{\vartheta \in T} d_{\text{TV}}(X_n, X_n^\vartheta) \leq \sum_{r=1}^K \mathbb{P}(I_r^{(n)} = n) \sup_{\vartheta \in T} d_{\text{TV}}(X_n, X_n^\vartheta). \quad (15)$$

Here, we use the total variation distance $d_{\text{TV}}(X_n, X_n^\vartheta) := d_{\text{TV}}(\mathbb{P}_{X_n}, \mathbb{P}_{X_n^\vartheta})$, where

$$d_{\text{TV}}(\mu, \nu) = \sup_{A \subseteq D[0,1] \text{ measurable}} |\mu(A) - \nu(A)|$$

for probability measures μ, ν on $D[0,1]$. As $\sum_{r=1}^K \mathbb{P}(I_r^{(n)} = n) < 1$ for all $n \geq 2$, it follows from (15) that $X_n \stackrel{d}{=} X_n^\vartheta$ for all $\vartheta \in T$ finishing the inductive proof.

Convergence of the discrete process. We continue with the definition of an accompanying sequence Q_n^ϑ by replacing the coefficients in the discrete recurrence by their limits. With $Q_0^\vartheta = Q_1^\vartheta := 0$ for all $\vartheta \in T$ and convention (9), let

$$Q_n^\vartheta = \sum_{r=1}^b \frac{1}{\sqrt{b}} \mathfrak{A}_{\frac{F_{r-1}^{n,\vartheta}}{n}, \frac{F_r^{n,\vartheta}}{n}}(Q_{I_r^{n,\vartheta}}^{\vartheta r}) + \sum_{r=1}^b \mathbf{1}_{[\frac{F_{r-1}^{n,\vartheta}}{n}, \frac{F_r^{n,\vartheta}}{n})} \mathcal{N}_r^\vartheta.$$

Again, on the event M , this defines a family of càdlàg processes. By arguments similar to the verification of $X_n^\vartheta \stackrel{d}{=} X_n$ for all $n \geq 2, \vartheta \in T$ above, one can show that, for all $n \geq 2$, the distributions of Q_n^ϑ and $X_n^\vartheta - Q_n^\vartheta$ do not depend on ϑ . The proof of $\Delta_n := \mathbb{E}[\|X_n^\vartheta - Q_n^\vartheta\|^3] \rightarrow 0$ is standard in the context of the contraction method. We have

$$\begin{aligned} \Delta_n &\leq \sum_{r=1}^b \mathbb{E} \left[\left\| \sqrt{\frac{I_r^{n,\vartheta}}{n}} \mathfrak{A}_{\frac{F_{r-1}^{n,\vartheta}}{n}, \frac{F_r^{n,\vartheta}}{n}}(X_{I_r^{n,\vartheta}}^{\vartheta r} - Q_{I_r^{n,\vartheta}}^{\vartheta r}) \right. \right. \\ &\quad \left. \left. + \left(\sqrt{\frac{I_r^{n,\vartheta}}{n}} - \frac{1}{\sqrt{b}} \right) \mathfrak{A}_{\frac{F_{r-1}^{n,\vartheta}}{n}, \frac{F_r^{n,\vartheta}}{n}}(Q_{I_r^{n,\vartheta}}^{\vartheta r}) \right\|^3 \right] \end{aligned} \quad (16)$$

$$+ \mathbf{1}_{[\frac{F_{r-1}^{n,\vartheta}}{n}, \frac{F_r^{n,\vartheta}}{n})} \left(\frac{b}{b-1} \frac{I_r^{n,\vartheta} - \frac{1}{b}n}{\sqrt{n}} - \mathcal{N}_r^\vartheta \right) \left\| \right\|^3. \quad (17)$$

By construction, $\|Q_n^\vartheta\| \leq \sum_{n \geq 0} \|G_{n+1}^\vartheta - G_n^\vartheta\|$ showing that all moments of the sequence $\|Q_n^\vartheta\|$ are bounded. As $\sqrt{I_r^{n,\vartheta}/n} \rightarrow b^{-1/2}$ with respect to all moments, it follows that all moments of the supremum over $t \in [0,1]$ of the summand in (16) vanish as $n \rightarrow \infty$. The same is true for the term in (17) since the almost sure convergence in (11) gives rise to convergence of arbitrary moments. Thus, upon applying Hölder's inequality to deal with mixed moments, there exists a sequence $\varepsilon(n) \rightarrow 0$,

such that,

$$\begin{aligned}\Delta_n &\leq (1 + \varepsilon(n)) \sum_{r=1}^b \mathbb{E} \left[\left\| \sqrt{\frac{I_r^{n,\vartheta}}{n}} \mathfrak{A}_{\frac{F_r^{n,\vartheta}}{n}, \frac{F_r^{n,\vartheta}}{n}} \left(X_{I_r^{n,\vartheta}}^{\vartheta r} - Q_{I_r^{n,\vartheta}}^{\vartheta r} \right) \right\|^3 \right] \\ &\leq (1 + \varepsilon(n)) \sum_{r=1}^b \mathbb{E} \left[\left\| \sqrt{\frac{I_r^{n,\vartheta}}{n}} \left(X_{I_r^{n,\vartheta}}^{\vartheta r} - Q_{I_r^{n,\vartheta}}^{\vartheta r} \right) \right\|^3 \right].\end{aligned}$$

From here, using $\mathbb{E} \left[\left(\frac{I_r^{n,\vartheta}}{n} \right)^{3/2} \right] \rightarrow b^{-3/2} < 1$, a simple induction on n shows that Δ_n is bounded.

In a second step, by the same contractive argument, one can show that $\Delta_n \rightarrow 0$. We omit the details which are standard in the framework of the contraction method.

Finally, the convergence $Q_n^\vartheta \rightarrow G^\vartheta$ in probability with respect to the Skorokhod topology is obtained by aligning the points of discontinuity of Q_n^ϑ to those of G_n^ϑ . This can be worked out in the same way as in [18], see the proof of Proposition 3.5 there. \square

Proof of Proposition 2.2. Assume X (more precisely, its distribution) solves (10). Then, we have

$$X(0) \stackrel{d}{=} \frac{1}{\sqrt{b}} X(0) + \mathcal{N}_1, \quad (18)$$

where $X(0)$ and \mathcal{N}_1 are independent and \mathcal{N}_1 has the distribution $\mathcal{N}(0, (b-1)^{-2})$. Classical results from the theory of perpetuities, see, e.g. Theorem 1.5 in [20], show that this identity uniquely determines the distribution of $X(0)$. Thus, $X(0)$ has the normal distribution $\mathcal{N}(0, b/(b-1)^2) =: \mu$. By the same argument, $X(1)$ has distribution μ . Next, for $k = 1, \dots, b-1$, we obtain

$$X(kb^{-1}) \stackrel{d}{=} \frac{1}{\sqrt{b}} X(0) + \mathcal{N}_1,$$

with conditions as in (18). Thus, for all $k = 1, \dots, b-1$, $X(kb^{-1})$ has distribution μ . Inductively, one shows that $X(t)$ has distribution μ for any b -adic t . By right continuity, the same holds for all $t \in [0, 1]$.

Next, let $n \geq 2, \leq t_1 < t_2 < \dots < t_n \leq 1$ and $j^* = \max\{j(t_k, t_{k+1}) : 1 \leq k \leq n-1\}$. Denote by ε_i^j the j -th digit of the b -ary expansion of t_i and $\vartheta_i^j = \varepsilon_i^1 \varepsilon_i^2 \dots \varepsilon_i^j$ for $j \geq 1$. We abbreviate $\vartheta_i^0 := \emptyset$. Let $\{\mathcal{N}^\vartheta, \vartheta \in \tilde{T}, |\vartheta| \leq j^* - 1\}$ be a family of independent random variables where each $\mathcal{N}^\vartheta = (\mathcal{N}_0^\vartheta, \dots, \mathcal{N}_{b-1}^\vartheta)$ is distributed like in (6). Here, \tilde{T} denotes the b -ary tree based on the digits $\{0, \dots, b-1\}$. Independently of this family, let X^1, \dots, X^n be independent copies of X . Then, developing (10) for j^* levels, there exist deterministic $0 \leq s_1, \dots, s_n \leq 1$ such that, for any $\alpha_1, \dots, \alpha_n \in \mathbb{R}$, we have

$$\sum_{i=1}^n \alpha_i X(t_i) \stackrel{d}{=} b^{-j^*/2} \sum_{i=1}^n \alpha_i X^i(s_i) + \sum_{i=1}^n \alpha_i \sum_{k=0}^{j^*-1} b^{-j/2} \mathcal{N}_{\varepsilon_i^{k+1}}^{\vartheta_i^k}.$$

By construction, the right hand side of this display has a normal distribution with zero mean. Thus, X is a Gaussian process. The covariance function is uniquely determined by (10) which proves the claim. \square

The last proof relied on the fact that the marginal distributions of any solution of (10) are Gaussian. An alternative proof which also extends to non-Gaussian solutions to fixed-point equation of type (10), where for each $t \in [0, 1]$, only of the processes X^1, \dots, X^b on the right hand side contributes, can be given following the arguments applied in the proof of Lemma 4.2 in [7].

2.2 The Markov source model

In this and all remaining paragraphs of Section 2, we study the Radix Selection algorithm in the Markov source model with restriction to $b = 2$ buckets. The analysis is considerably more involved as in the last section; in particular, a result analogous to Theorem 2.1 does not hold in the most general model. The main results of the section are presented in Section 2.2.2: we prove a law of large numbers and an expansion of the mean for $Y_n^\mu(\lfloor tn \rfloor + 1), t \in [0, 1]$, under the general set-up. We also state a functional central limit theorem comparable to Theorem 2.1 in the asymmetric case when $p_{00} = p_{10}$. We start by introducing a family of càdlàg functions m_μ which play an important role subsequently. Throughout the section, we abbreviate

$$p_+ := p_{00} \vee p_{01} \vee p_{10} \vee p_{11}, \quad (19)$$

and recall $p_+ < 1$.

2.2.1 A family of limiting functions m_μ

For $k \geq 1$ and $i = 0, 1$ we recursively define sets $\mathcal{D}_k^i = \{s_{k,\ell}^i \mid \ell = 0, \dots, 2^k\}$ as follows: For $k = 1$ we set $(s_{1,0}^i, s_{1,1}^i, s_{1,2}^i) := (0, p_{i0}, 1)$ for $i = 0, 1$. Further, for all $k \geq 1$, $i = 0, 1$ and $0 \leq \ell \leq 2^k$ we set

$$s_{k+1,\ell}^i := \begin{cases} s_{k,\ell/2}^i, & \text{if } \ell \bmod 4 \in \{0, 2\}, \\ p_{00}s_{k,(\ell+1)/2}^i + p_{01}s_{k,(\ell-1)/2}^i, & \text{if } \ell \bmod 4 = 1. \\ p_{10}s_{k,(\ell+1)/2}^i + p_{11}s_{k,(\ell-1)/2}^i, & \text{if } \ell \bmod 4 = 3, \end{cases}$$

We further define $\mathcal{D}_\infty^i := \cup_{k=1}^\infty \mathcal{D}_k^i$, $\mathcal{C}_k^i = \mathcal{D}_k^i \setminus \{0, 1\}$, $k \geq 1$, and $\mathcal{C}_\infty^i := \cup_{k=1}^\infty \mathcal{C}_k^i$. Note that, for each $k \geq 1$, the set \mathcal{D}_k^i decomposes the unit interval into 2^k sub-intervals. For $t \notin \mathcal{C}_k^i$ we denote by $\lambda_k^i(t)$ the length of the (unique) sub-interval of this decomposition that contains t . Finally, let $\ell_k^i(t)$ be the unique value ℓ such that $\lambda_k^i(t) = s_{k,\ell+1}^i - s_{k,\ell}^i$. Moreover, we set $\varepsilon_0^i(t) = i$ and, for $k \geq 1$, $\varepsilon_k^i(t) = 0$ if $\ell_k^i(t)$ is even and $\varepsilon_k^i(t) = 1$ otherwise.

Next, for $i = 0, 1$ and $t \notin \mathcal{C}_\infty^i$, abbreviating $\lambda_0^i(\cdot) \equiv 1$, we set

$$m_i(t) := \sum_{n=0}^\infty \lambda_n^i(t). \quad (20)$$

For $t \in [0, 1]$, let $r^i(t) \geq 1$ be minimal with $t \in \mathcal{C}_{r^i(t)}^i$ abbreviating $r^i(t) = \infty$ for $t \notin \mathcal{C}_\infty^i$. Then, for $t \in \mathcal{C}_\infty^i$, let

$$m_i(t) := \sum_{j=0}^{r^i(t)-1} \lambda_j^i(t) + \frac{1}{2} \lambda_{r^i(t)-1}^i(t) \left(1 + p_{\varepsilon_{r^i(t)-1}} \frac{p_{01}}{p_{10}} + p_{\varepsilon_{r^i(t)-1}} \frac{p_{10}}{p_{01}} \right), \quad (21)$$

and note that this definition is consistent with (20) for $t \notin \mathcal{C}_\infty^i$.

Further, for an initial distribution $\mu = \mu_0 \delta_0 + \mu_1 \delta_1$ with $\mu_0 \in [0, 1]$, we denote

$$\mathcal{D}_\infty^\mu := \mu_0 \mathcal{D}_\infty^0 \cup (\mu_0 + \mu_1 \mathcal{D}_\infty^1), \quad \mathcal{C}_\infty^\mu := \mathcal{D}_\infty^\mu \setminus \{0, 1\},$$

and, for $t \notin \mathcal{C}_\infty^\mu$,

$$m_\mu(t) := \begin{cases} \mu_0 m_0\left(\frac{t}{\mu_0}\right) + 1, & \text{if } t < \mu_0, \\ (1 - \mu_0) m_1\left(\frac{t - \mu_0}{1 - \mu_0}\right) + 1, & \text{if } t > \mu_0. \end{cases} \quad (22)$$

The functions m_0 and m_1 have the following properties:

i) m_0 and m_1 are bounded on $[0, 1]$.

ii) If $p_{00} = p_{10} = 1 - p$ then

$$m_0(t) = m_1(t) = \frac{2p-1}{p(1-p)}t + \frac{1}{p}.$$

iii) If $p_{00} \neq p_{10}$ then m_i is continuous at t if and only if $t \notin \mathcal{C}_\infty^i$. At points of discontinuity, left and right limits exist and we have

$$m_i(t-) = \sum_{j=0}^{r^i(t)-1} \lambda_j^i(t) + \lambda_{r^i(t)-1}^i(t) p_{\varepsilon_{r^i(t)-1}^i 0} \left(1 + \frac{p_{01}}{p_{10}}\right),$$

$$m_i(t+) = \sum_{j=0}^{r^i(t)-1} \lambda_j^i(t) + \lambda_{r^i(t)-1}^i(t) p_{\varepsilon_{r^i(t)-1}^i 1} \left(1 + \frac{p_{10}}{p_{01}}\right).$$

Thus, $m_i(t) = \frac{1}{2}(m_i(t-) + m_i(t+))$.

iv) For $i = 0, 1$ and $t \in [0, 1] \setminus \{p_{i0}\}$,

$$m_i(t) = \mathbf{1}_{[0, p_{i0})}(t) p_{i0} m_0 \left(\frac{t}{p_{i0}} \right) + \mathbf{1}_{(p_{i0}, 1]}(t) p_{i1} m_1 \left(\frac{t - p_{i0}}{p_{i1}} \right) + 1.$$

i) follows since $\lambda_{k,\ell} \leq p_+^k$. To prove ii), first note that $m_0 = m_1$. Assume $f, g : [0, 1] \rightarrow \mathbb{R}$ were two bounded solutions to the fixed-point equation in iv). Then, $\|f - g\| \leq p_+ \|f - g\|$, hence $f = g$. We deduce that there is at most one bounded solution of the fixed-point equation. Thus, ii) follows from verifying that the term in ii) satisfies the fixed-point equation which is elementary. Next, note that, for $t < p_{i0}$, we have $m_i(t) = p_{i0} m_0(t/p_{i0}) + 1$ since $\lambda_{k+1}^i(t) = p_{i0} \lambda_k^0(t/p_{i0})$ for $k \geq 0$. The analogous statement for $t > p_{i0}$ proves iv). iii) follows easily from the definitions of m_0, m_1 .

2.2.2 Main results in the Markov source model

We have the following asymptotic behavior of the average complexity:

Theorem 2.3. *Let $Y_n^\mu(\ell)$ denote the number of bucket operations of Radix Selection with $b = 2$ selecting a rank $1 \leq \ell \leq n$ among n independent data generated from the Markov source model with initial distribution $\mu = \mu_0 \delta_0 + \mu_1 \delta_1$ where $\mu_0 \in [0, 1]$ and transition matrix $(p_{ij})_{i,j \in \{0,1\}}$ with $p_+ < 1$. We abbreviate $Y_n^i = Y_n^{\mu^i}(n)$ for $i = 1, 2$ where $\mu_0^0 = p_{00}$ and $\mu_0^1 = p_{10}$.*

For all $t \in [0, 1]$, we have

$$n^{-1} \mathbb{E}[Y_n^i(\lfloor tn \rfloor + 1)] \rightarrow m_i(t), \quad (23)$$

with convergence in L^1 if and only if $t \notin \mathcal{C}_\infty^i$.

For all $t \notin \mathcal{C}_\infty^\mu$, with convergence in L^1 , we have

$$n^{-1} Y_n^\mu(\lfloor tn \rfloor + 1) \rightarrow m_\mu(t). \quad (24)$$

If $t_n \rightarrow t \notin \mathcal{C}_\infty^\mu$, then, with convergence in L^1 ,

$$n^{-1} Y_n^\mu(\lfloor t_n n \rfloor + 1) \rightarrow m_\mu(t). \quad (25)$$

The distributional behavior of the normalized sequences

$$X_n^i(t) = \frac{Y_n^i(\lfloor tn + 1 \rfloor) - m_i(t)n}{\sqrt{n}}, \quad t \in [0, 1], i = 0, 1,$$

is more involved; in particular, there is no functional convergence in the Skorokhod space unless $p_{00} = p_{10}$.

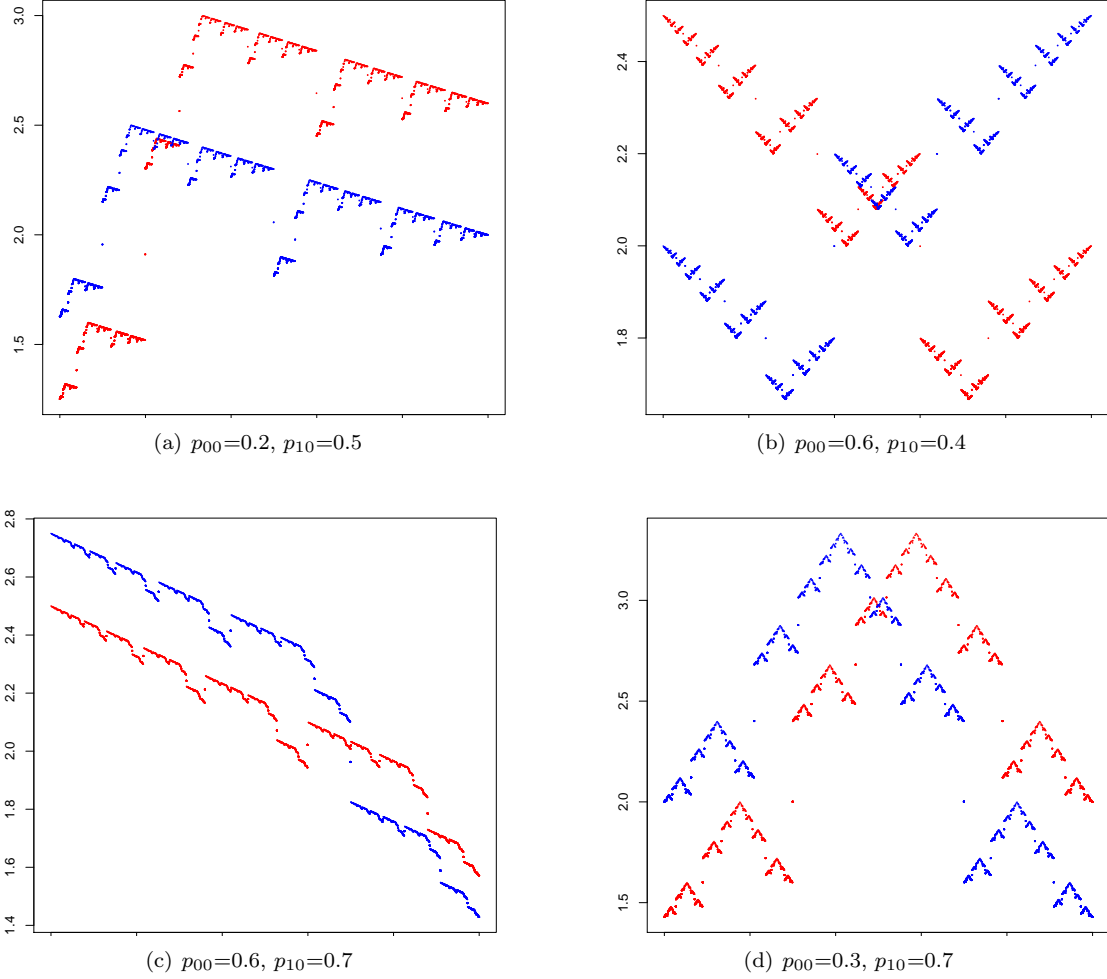


Figure 2: Plots of m_0 (red) and m_1 (blue) for different Markov sources.

Theorem 2.4. *Let $p_{00} \neq p_{10}$. Then, both sequences $(\|X_n^0\|)_{n \geq 1}$ and $(\|X_n^1\|)_{n \geq 1}$ are not tight.*

The proof of Theorem 2.4 presented below relies on the discontinuities of the mean functions m_0, m_1 when $p_{00} \neq p_{10}$. Indeed, for $p_{00} = p_{10}$, we encounter the so-called *asymmetric Bernoulli model* in which the mean function m corresponding to m_0, m_1, m_μ in Theorem 2.3 becomes an affine function. Here, we obtain a functional limit law as for the uniform model in Theorem 2.1.

Theorem 2.5. *Consider Radix Selection using $b = 2$ buckets on a set of n independent data generated from the asymmetric Bernoulli model with success probability $p := p_{00} = p_{10} \in (0, 1) \setminus \{\frac{1}{2}\}$. For the process $X_n^{\text{asyB}} = (X_n^{\text{asyB}}(t))_{0 \leq t \leq 1}$ of the normalized number of bucket operations $Y_n^{\text{asyB}}(\ell)$ defined by*

$$X_n^{\text{asyB}}(t) := \frac{Y_n^{\text{asyB}}(\lfloor tn \rfloor + 1) - m(t)n}{\sqrt{n}}, \quad t \in [0, 1],$$

with $Y_n^{\text{asyB}}(n+1) := Y_n^{\text{asyB}}(n)$ and

$$m(t) = \frac{2p-1}{p(1-p)}t + \frac{1}{p}, \quad t \in [0, 1], \quad (26)$$

we have weak convergence, as $n \rightarrow \infty$, in $(D[0, 1], d_{sk})$:

$$X_n^{\text{asyB}} \xrightarrow{d} G^{\text{asyB}}.$$

Here, $G^{\text{asyB}} = (G^{\text{asyB}}(t))_{t \in [0, 1]}$ is a centered Gaussian process with covariance function

$$\mathbb{E}[G^{\text{asyB}}(s)G^{\text{asyB}}(t)] = - \prod_{k=1}^{r(s,t)} p[g(t, k)] + \sum_{k=1}^{r(s,t)} \frac{\prod_{j=1}^k p[g(t, j)]}{p[1 - g(t, k)]}, \quad s, t \in [0, 1],$$

where $p[0] := 1 - p$, $p[1] := p$ and the functions $r : [0, 1]^2 \rightarrow \mathbb{N}_0 \cup \{\infty\}$ and $g : [0, 1] \times \mathbb{N}_0 \rightarrow \{0, 1\}$ are defined as follows:

$$r(s, t) = \max\{n \in \mathbb{N}_0 | g(s, \ell) = g(t, \ell), 1 \leq \ell \leq n\}$$

and g and $h : [0, 1] \times \mathbb{N}_0 \rightarrow [0, 1]$ are recursively defined by $g(t, 0) = 0$, $h(t, 0) = t$ for $t \in [0, 1]$ and for $k \geq 1$ by

$$g(t, k) = \begin{cases} 0, & \text{if } h(t, k-1) < 1-p, \\ 1, & \text{if } h(t, k-1) \geq 1-p, \end{cases}$$

$$h(t, k) = \begin{cases} \frac{h(t, k-1)}{1-p}, & \text{if } h(t, k-1) < 1-p, \\ \frac{h(t, k-1) - (1-p)}{p}, & \text{if } h(t, k-1) \geq 1-p. \end{cases}$$

We leave it as an open problem to decide whether the one- (or finite-) dimensional marginal distributions of the processes $(X_n^0)_{n \geq 1}, (X_n^1)_{n \geq 1}$ converge weakly in the general case $p_{00} \neq p_{10}$.

All results in this section rely on the distributional recurrence for the complexity Y_n^i similar to (5): for $n \geq 2$, the design of the algorithm yields

$$Y_n^i \stackrel{d}{=} \left(\mathbf{1}_{[1, I_n^i]}(\ell) Y_{I_n^i}^0(\ell) + \mathbf{1}_{(I_n^i, n]}(\ell) Y_{n-I_n^i}^1(\ell - I_n^i) + n \right)_{1 \leq \ell \leq n}, \quad i = 0, 1, \quad (27)$$

where $Y_0^0, \dots, Y_n^0, Y_0^1, \dots, Y_n^1, I_n^0, I_n^1$ are independent (the independence between I_n^0 and I_n^1 is not required) and I_n^i is $B(n, p_{i0})$ distributed for $i = 0, 1$. Moreover, for general initial distribution μ , we have

$$Y_n^\mu \stackrel{d}{=} \left(\mathbf{1}_{[1, K_n]}(\ell) Y_{K_n}^0(\ell) + \mathbf{1}_{(K_n, n]}(\ell) Y_{n-K_n}^1(\ell - K_n) + n \right)_{1 \leq \ell \leq n}, \quad (28)$$

where $Y_0^0, \dots, Y_n^0, Y_0^1, \dots, Y_n^1, K_n$ are independent and K_n has the binomial $B(n, \mu_0)$ distribution.

2.2.3 Proofs of Theorems 2.3, 2.4 and 2.5

In addition to the distributional recurrences (27) and (28), our approach towards Theorem 2.3 uses a direct construction of a version of the random variables $Z_n^i(t) := Y_n^i(\lfloor tn \rfloor + 1)$ where $Y_n^i(n+1) := Y_n^i(n)$. To this end, let $\{J_{k,\ell}^0 : k \geq 1, 0 \leq \ell \leq 2^{k-1} - 1\}$, $\{J_{k,\ell}^1 : k \geq 1, 0 \leq \ell \leq 2^{k-1} - 1\}$ be two independent sets of independent families of random variables where $J_{k,\ell}^i(n)$ has the binomial distribution $B(n, p_{i0})$.

Analogously to the sequence of sets $(\mathcal{D}_n^i)_{n \geq 0}$, we now define a nested decomposition of the set $\{0, \dots, n\}$. We set $(I_{1,0}^i(n), I_{1,1}^i(n), I_{1,2}^i(n)) := (0, J_{1,0}^i(n), n)$ for $i = 0, 1$. Further, for all $k \geq 1$, $i = 0, 1$ and $0 \leq \ell \leq 2^k$ we set

$$I_{k+1,\ell}^i(n) := \begin{cases} I_{k,\ell/2}^i(n), & \text{if } \ell \bmod 4 \in \{0, 2\}, \\ I_{k,(\ell-1)/2}^i + J_{k+1,(\ell-1)/2}^0(I_{k,(\ell+1)/2}^i - I_{k,(\ell-1)/2}^i), & \text{if } \ell \bmod 4 = 1, \\ I_{k,(\ell-1)/2}^i + J_{k+1,(\ell-1)/2}^1(I_{k,(\ell+1)/2}^i - I_{k,(\ell-1)/2}^i), & \text{if } \ell \bmod 4 = 3, \end{cases}$$

The vector $(I_{k,\ell+1}^i(n) - I_{k,\ell}^i(n))_{0 \leq \ell \leq 2^k-1}$ is multinomially distributed with parameters $n; s_{k,1}^i - s_{k,0}^i, \dots, s_{k,2^k}^i - s_{k,2^k-1}^i$. For $t \in [0, 1 - n^{-1}]$ let $L_k^i(t, n)$ be the unique value ℓ with $I_{k,\ell}^i(n) \leq tn < I_{k,\ell+1}^i(n)$. For $t > 1 - n^{-1}$, let $L_k^i(t, n)$ be the largest value ℓ with $I_{k,\ell}^i(n) < n$. For all $t \in [0, 1]$, define $\Lambda_k^i(t, n) = I_{k,L_k^i(t,n)+1}^i(n) - I_{k,L_k^i(t,n)}^i(n)$. Finally, let $E_k^i(t, n) = 0$ if $L_k^i(t, n)$ is even and $E_k^i(t, n) = 1$ otherwise. We abbreviate $\Lambda_0^i(\cdot, n) \equiv n$ for any $n \geq 0$ and $x_+ = x\mathbf{1}_{(1,\infty)}(x)$ for $x \geq 0$. Note that $\Lambda_k^i(t, n)$ is distributed like the number of strings that share a common prefix of length k with the element or rank $\lfloor tn \rfloor + 1$ among n independent strings generated by a Markov source with initial distribution μ^i . Developing the recurrence leads to

$$Z_n^i(t) \stackrel{d}{=} \sum_{j=0}^{k-1} \Lambda_j^i(t, n)_+ + Z_{\Lambda_k^i(t,n)}^{E_k^i(t,n)}(h_k^i(t, n)) \quad (29)$$

for some value $0 \leq h_k^i(t, n) \leq 1$ where $(E_k^i(t, n), h_k^i(t, n), \Lambda_k^i(t, n))$ and $(Z_n^0)_{n \geq 0}, (Z_n^1)_{n \geq 0}$ are independent. Note that $h_k^i(t, n)$ is a function of $(I_{k,L_k^i(t,n)}^i(n), I_{k,L_k^i(t,n)+1}^i(n))$.

Next, for $t \in [0, 1]$ and $k \leq r_i(t) - 1$, let $\Delta_k^i(t, n) = I_{k,\ell_k(t)+1}^i(n) - I_{k,\ell_k(t)}^i(n)$ which has the binomial distribution $B(n, \lambda_k^i(t))$. Set $R_{k,n}^i(t) = \sum_{j=0}^k \Lambda_j^i(t, n)_+ - \sum_{j=1}^k \Delta_j^i(t, n)$. Then,

$$Z_n^i(t) \stackrel{d}{=} \sum_{j=0}^k \Delta_j^i(t, n) + R_{k,n}^i(t) + Z_{\Lambda_{k+1}^i(t,n)}^{E_{k+1}^i(t,n)}(h_{k+1}^i(t, n)).$$

The proof of Theorem 2.3 and other results in Sections 2 and 3 use Theorem 4.5 from Section 4.2. Note that the proof of Theorem 4.5 is self-contained.

Proof of Theorem 2.3. Let $t \notin \mathcal{D}_\infty^i$. With $\Delta_0^i(t, n) = n$, for all $k \geq 0$, we have

$$\frac{Z_n^i(t)}{n} - m_i(t) \stackrel{d}{=} \underbrace{n^{-1} \sum_{j=0}^k \Delta_j^i(t, n) - m_i(t)}_{A_{k,n}} + n^{-1} \left(R_{k,n}^i(t) + Z_{\Lambda_{k+1}^i(t,n)}^{E_{k+1}^i(t,n)}(h_{k+1}^i(t, n)) \right).$$

Let $\varepsilon > 0$. In order to bound the term $\mathbb{E}[|A_{k,n}|]$ note that, by Jensen's inequality, we have

$$\mathbb{E}[|\text{Bin}(n, p) - np|] \leq \sqrt{\text{Var}(\text{Bin}(n, p))} \leq \sqrt{n}/2.$$

By an adaption of the triangle inequality and the last bound, we deduce

$$\mathbb{E}[|A_{k,n}|] \leq kn^{-1/2}/2 + p_+^{k+1}(1 - p_+)^{-1}.$$

Thus, fixing k sufficiently large, for all n large enough, we have $\mathbb{E}[|A_{k,n}|] \leq \varepsilon$. Next, as a consequence of the convergence given in Theorem 4.5, there exists a constant $C > 0$ satisfying $\mathbb{E}[Y_n^i(\ell)] \leq Cn$ for all $n \geq 1, i = 1, 2$ and $1 \leq \ell \leq n$. Therefore,

$$\begin{aligned} & n^{-1} \mathbb{E} \left[Z_{\Lambda_{k+1}^i(t,n)}^{E_{k+1}^i(t,n)}(h_{k+1}^i(t, n)) \right] \\ & \leq Cn^{-1} \mathbb{E} [\Lambda_{k+1}^i(t, n)] \\ & \leq Cn^{-1} \mathbb{E} \left[\max_{0 \leq \ell \leq 2^k-1} I_{k+1,\ell+1}^i(n) - I_{k+1,\ell}^i(n) \right]. \end{aligned} \quad (30)$$

By the theorem of dominated convergence and the strong law of large numbers,

$$\limsup_{n \rightarrow \infty} \sup_{t \in [0,1]} n^{-1} \mathbb{E} \left[Z_{\Lambda_{k+1}^i(t,n)}^{E_{k+1}^i(t,n)}(h_{k+1}^i(t, n)) \right] \leq Cp_+^{k+1}. \quad (31)$$

Thus, upon fixing k sufficiently large, the left hand side of (30) is bounded from above by ε for all n large enough. Finally, note that

$$R_{k,n}^i(t) = \sum_{j=1}^k [\Lambda_j^i(t, n)_+ - \Delta_j^i(t, n)] \mathbf{1}_{\{\ell_j^i(t) \neq L_j^i(t, n)\}} - \#\{1 \leq j \leq k : \Delta_j^i(t, n) = 1\}.$$

We have $\ell_k^i(t) = L_k^i(t, n)$ if $I_{k, \ell_k^i(t)}^i(n)/n$ and $s_{k, \ell_k^i(t)}^i$ are sufficiently close. $I_{k, \ell_k^i(t)}^i(n)$ has the binomial distribution $B\left(n, s_{k, \ell_k^i(t)}^i\right)$ where $0 < s_{k, \ell_k^i(t)}^i < 1$. Let $\alpha = 2 \min \left\{ t - s_{k, \ell_k^i(t)}^i, s_{k, \ell_k^i(t)+1}^i - t \right\}$. If

$$|I_{k, m}^i(n)/n - s_{k, m}^i| < \frac{\alpha}{2} \lambda_k^i(t) \quad (32)$$

both for $m = \ell_k^i(t)$ and $m = \ell_k^i(t) + 1$, then $\ell_k^i(t) = L_k^i(t, n)$. In particular, $\ell_j^i(t) = L_j^i(t, n)$ for all $1 \leq j \leq k$. Since,

$$\mathbb{E} \left[|\Delta_j^i(t, n) - \Lambda_j^i(t, n)_+| \mathbf{1}_{\{\ell_j^i(t) \neq L_j^i(t, n)\}} \right] \leq 2n\mathbb{P}(\ell_j^i(t) \neq L_j^i(t, n)), \quad (33)$$

it follows that the left hand side of the last display tends to zero exponentially fast as $n \rightarrow \infty$ upon using Chernoff's inequality to bound the probabilities of the complements of the events (32) for $m = \ell_k^i(t)$ and $m = \ell_k^i(t) + 1$. Summarizing, for any fixed $k \geq 1$, we have $\mathbb{E}[|R_n^i(t)|] \leq \varepsilon n + k$ for all n sufficiently large finishing the proof of (23) for $t \notin \mathcal{D}_\infty^i$.

For $t \notin \mathcal{D}_\infty^i$ and $t_n \rightarrow t$, the same arguments can be used to infer (24) relying on the uniformity in $t \in [0, 1]$ in the bound (31). Our proof can easily be modified to prove (23) and (24) for the boundary cases $t \in \{0, 1\}$ noting that $\ell_k^i(0) \neq L_k^i(0, n)$ only if $I_{k, 1}^i(n) = 0$ which happens with exponentially small probability (analogously for $t = 1$).

In order to prove (23) for $t \in \mathcal{C}_\infty^i$, recall the development

$$Z_n^i(t) \stackrel{d}{=} \sum_{j=0}^{r^i(t)-1} \Lambda_j^i(t, n)_+ + Z_{\Lambda_{r^i(t)}^i(t, n)}^{E_{r^i(t)}^i(t, n)}(h_{r^i(t)}^i(t, n))$$

with conditions as in (29). As in the proof above, one can show that, with convergence in L^1 ,

$$n^{-1} \sum_{j=0}^{r^i(t)-1} \Lambda_j^i(t, n)_+ \rightarrow \sum_{j=0}^{r^i(t)-1} \lambda_j^i(t). \quad (34)$$

Abbreviate $r = r^i(t)$ and $\gamma_n^i(t) := \mathbb{E}[Z_n^i(t)] = \alpha_n^i(\lfloor nt \rfloor + 1)$. We have

$$\begin{aligned} \mathbb{E} \left[Z_{\Lambda_r^i(t, n)}^{E_r^i(t, n)}(h_r^i(t, n)) \right] &= \mathbb{E} \left[\gamma_{\Lambda_r^i(t, n)}^{E_r^i(t, n)}(h_r^i(t, n)) \right] \\ &= \mathbb{E} \left[\gamma_{\Lambda_r^i(t, n)}^{E_r^i(t, n)}(h_r^i(t, n)) \mathbf{1}_{\{\ell_{r-1}^i(t) = L_{r-1}^i(t, n)\}} \right] + \mathbb{E} \left[\gamma_{\Lambda_{r-1}^i(t, n)}^{E_r^i(t, n)}(h_r^i(t, n)) \mathbf{1}_{\{\ell_{r-1}^i(t) \neq L_{r-1}^i(t, n)\}} \right]. \end{aligned}$$

Using Theorem 4.5, the second summand on the right hand side is bounded from above by $Cn\mathbb{P}(\ell_{r-1}^i(t) \neq L_{r-1}^i(t, n))$ with C as in (30). Hence, it vanishes exponentially fast similarly to the term in (33).

Let $\ell := 2\ell_{r-1}^i(t) + 1$ such that $t = s_{r, \ell}^i$ and suppress the dependency on n and i in the quantities $I_{k, \ell}^i(n)$ to simplify the notation. Then, the first summand on the right hand side of the last display equals

$$\begin{aligned} &\mathbb{E} \left[\gamma_{I_{r, \ell} - I_{r, \ell-1}}^0 \left(\frac{tn - I_{r, \ell-1}}{I_{r, \ell} - I_{r, \ell-1}} \right) \mathbf{1}_{\{I_{r, \ell-1} \leq tn < I_{r, \ell}\}} \right] \\ &+ \mathbb{E} \left[\gamma_{I_{r, \ell+1} - I_{r, \ell}}^1 \left(\frac{tn - I_{r, \ell}}{I_{r, \ell+1} - I_{r, \ell}} \right) \mathbf{1}_{\{I_{r, \ell} \leq tn < I_{r, \ell+1}\}} \right]. \end{aligned} \quad (35)$$

The random variables in the last display depend only on $(I_{r,\ell-1}, I_{r,\ell}, I_{r,\ell+1})$. For simplicity, we may assume that the vector $(I_{r,\ell-1}, I_{r,\ell}, I_{r,\ell+1})$ converges almost surely after rescaling; in particular, $(I_{r,\ell} - nt)/\sqrt{n} \rightarrow \mathcal{N}$ for a random variable \mathcal{N} having the normal distribution $\mathcal{N}(0, t(1-t))$. Obviously,

$$\liminf_{n \rightarrow \infty} \{I_{r,\ell-1} \leq tn < I_{r,\ell}\} = \{\mathcal{N} > 0\} \cup A,$$

with a \mathbb{P} -null set A . Almost surely, on the event in the last display,

$$\frac{tn - I_{r,\ell-1}}{I_{r,\ell} - I_{r,\ell-1}} \rightarrow 1, \quad n^{-1}(I_{r,\ell} - I_{r,\ell-1}) \rightarrow \lambda_{r-1}^i(t) p_{\varepsilon_{r-1}^i 0}.$$

In particular, using (25) for $t = 1$, almost surely,

$$n^{-1} \gamma_{I_{r,\ell} - I_{r,\ell-1}}^0 \left(\frac{tn - I_{r,\ell-1}}{I_{r,\ell} - I_{r,\ell-1}} \right) \mathbf{1}_{\{I_{r,\ell-1} \leq tn < I_{r,\ell}\}} \rightarrow \lambda_{r-1}^i(t) p_{\varepsilon_{r-1}^i 0} m_0(1) \mathbf{1}_{\mathcal{N} > 0}.$$

Using Theorem 4.5, the theorem of dominated convergence gives convergence of the mean in the last statement. The second term summand in (35) is handled analogously finishing the proof using (29), (34) and the definition of $m_i(t)$ in (21). Following the same argument on a distributional level, it is easy to see that there is no convergence in probability to a deterministic limit for $Z_{\Lambda_r^i(t,n)}^{E_r^i(t,n)}(h_r^i(t,n))$. Thus, there is no convergence in L^1 in (23).

Finally, (24) and (25) for general measures μ follow by similar arguments based on (28) again using the theorem of dominated convergence relying on (24), (25) for $\mu = \mu^0$ and $\mu = \mu^1$ and Theorem 4.5. We omit the details. \square

Proof of Theorem 2.4. For $i = 0, 1$, using the system of recurrences (27), we have

$$\begin{aligned} X_n^i &\stackrel{d}{=} \left(\sqrt{\frac{I_n^i}{n}} \mathfrak{A}_{0, \frac{I_n^i}{n}} \left(X_{I_n^i}^0 \right) (t) + \sqrt{\frac{n - I_n^i}{n}} \mathfrak{A}_{\frac{I_n^i}{n}, 1} \left(X_{n - I_n^i}^1 \right) (t) \right. \\ &\quad \left. + \frac{1}{\sqrt{n}} \left(n(1 - m_i(t)) + \mathfrak{A}_{0, \frac{I_n^i}{n}}(m_0)(t) I_n^i + \mathfrak{A}_{\frac{I_n^i}{n}, 1}(m_1)(t) (n - I_n^i) \right) \right)_{0 \leq t \leq 1}, \end{aligned}$$

with conditions on independence and identical distributions as in (27). We denote the additive term in the last display by $T_n^i(t), t \in [0, 1]$. Assume that $(\|X_n^0\|)_{n \geq 1}$ was tight. Then, for $\varepsilon > 0$, there exists $K > 0$ such that $\mathbb{P}(\|X_n^0\| > K) \leq \varepsilon$ for all $n \geq 0$. We have

$$\mathbb{P}(\|X_n^0\| \geq K) \geq \mathbb{P}(\|T_n^0 \mathbf{1}_{[0, I_n^0/n)}\| \geq 2K, \|X_{I_n^0}^0\| \leq K) \geq \mathbb{P}(\|T_n^0 \mathbf{1}_{[0, I_n^0/n)}\| \geq 2K) - \varepsilon \quad (36)$$

We investigate the behavior of the sequence $T_n := T_n^0$ abbreviating $p := p_{00}, I_n := I_n^0$ and $m := m_0$. Let $t \in \mathcal{C}_\infty^0$ with $t < p$ (e.g. $t = p^2$). Then, on the event $\{tn < I_n\}$, we have

$$\begin{aligned} T_n(t) &= \frac{1}{\sqrt{n}} \left(n - m(t)n + m \left(\frac{tn}{I_n} \right) I_n \right) \\ &= \frac{I_n - np}{\sqrt{n}} m \left(\frac{tn}{I_n} \right) + \sqrt{np} \left(m \left(\frac{t}{p} \right) - m \left(\frac{tn}{I_n} \right) \right). \end{aligned} \quad (37)$$

We may assume $(I_n - np)/\sqrt{n} \rightarrow \mathcal{N}$ almost surely, where \mathcal{N} has distribution $\mathcal{N}(0, p(1-p))$. Then, the first summand on the right hand side in (37) remains almost surely bounded as m is bounded on $[0, 1]$. Moreover, we have, almost surely,

$$\lim_{n \rightarrow \infty} m \left(\frac{tn}{I_n} \right) = \begin{cases} m \left(\frac{t}{p} \right), & \text{if } \mathcal{N} > 0, \\ m \left(\frac{t}{p} \right), & \text{if } \mathcal{N} < 0. \end{cases}$$

Since $t/p \in \mathcal{C}_\infty^0$, the function m has a double jump at t/p . It follows that $|T_n(t)|n^{-1/2}$ remains bounded away from zero almost surely. In particular, we have $\|T_n\| \rightarrow \infty$ almost surely. Using (36), this shows $\liminf_{n \rightarrow \infty} \mathbb{P}(\|X_n^0\| \geq K) \geq 1 - \varepsilon$ contradicting the tightness of $(\|X_n^0\|)_{n \geq 1}$. The analogous proof applies to the sequence $(\|X_n^1\|)_{n \geq 1}$. \square

Note that the proof can be generalized to show that, for any sequence $\alpha_n = o(n)$, the càdlàg processes $\alpha_n^{-1} (Y_n^i(\lfloor tn \rfloor + 1) - \mathbb{E}[Y_n^i(\lfloor tn \rfloor + 1)])$, $i = 0, 1$, are not tight.

We continue with the functional limit theorem in the asymmetric Bernoulli model when $p_{00} = p_{10}$. Since the proof of Theorem 2.5 is very similar to that of Theorem 2.1, we only sketch it here.

Sketch of the proof of Theorem 2.5. With $Y_n := Y_n^{\text{asyB}}$, we have

$$Y_n \stackrel{d}{=} (\mathbf{1}_{[1, I_n]}(\ell) Y_{I_n}^1(\ell) + \mathbf{1}_{(I_n, n]}(\ell) Y_{n-I_n}^2(\ell - I_n) + n)_{1 \leq \ell \leq n}, \quad (38)$$

where $(Y_j^1), (Y_j^2), I_n$ are independent and the Y_j^1, Y_j^2 have the same distribution as Y_j for all $j \geq 0$ and I_n has the binomial distribution $B(n, 1-p)$. Hence, for the process $X_n := X_n^{\text{asyB}}$, it follows

$$\begin{aligned} X_n \stackrel{d}{=} & \left(\sqrt{\frac{I_n}{n}} \mathfrak{A}_{0, \frac{I_n}{n}}(X_{I_n}^1)(t) + \sqrt{\frac{n-I_n}{n}} \mathfrak{A}_{\frac{I_n}{n}, 1}(X_{n-I_n}^2)(t) \right. \\ & \left. + \mathbf{1}_{[0, \frac{I_n}{n}]}(t) \frac{I_n - (1-p)n}{p\sqrt{n}} - \mathbf{1}_{[\frac{I_n}{n}, 1]}(t) \frac{I_n - (1-p)n}{(1-p)\sqrt{n}} \right)_{0 \leq t \leq 1} \end{aligned}$$

with conditions on independence and identical distributions as in the last display. In distribution, as $n \rightarrow \infty$,

$$\frac{I_n - (1-p)n}{\sqrt{n}} \rightarrow \mathcal{N},$$

where \mathcal{N} has distribution $\mathcal{N}(0, p(1-p))$. Thus, we arrive at the limiting equation

$$X \stackrel{d}{=} \sqrt{1-p} \mathfrak{A}_{0, 1-p}(X^1) + \sqrt{p} \mathfrak{A}_{1-p, 1}(X^2) + p^{-1} \mathbf{1}_{[0, 1-p]} \mathcal{N} - (1-p)^{-1} \mathbf{1}_{[1-p, 1]} \mathcal{N} \quad (39)$$

where X^1, X^2, \mathcal{N} are independent, X^1, X^2 have the distribution of X and \mathcal{N} has distribution $\mathcal{N}(0, p(1-p))$. The formal verification of the convergence $X_n \rightarrow X$ in distribution in the Skorokhod topology, where X is a centered Gaussian process with the given covariance structure satisfying the stochastic fixed-point equation (39), can be worked out as in the proof of Theorem 2.1. \square

3 The model of grand averages

We now consider the complexity of Radix Selection with $b = 2$ buckets assuming the Markov source model for the data and the model of grand averages for the rank. Let ξ be a random variable with the uniform distribution on $[0, 1]$ and independent of all remaining quantities.

Theorem 3.1. *Let $W_n^\mu = Y_n^\mu(\lfloor \xi n \rfloor + 1)$ denote the number of bucket operations of Radix Selection with $b = 2$ selecting the uniformly distributed rank $\lfloor \xi n \rfloor + 1$ from n independent data generated from the Markov source model with conditions as in Theorem 2.3. Then, as $n \rightarrow \infty$, in probability and with respect to convergence of all moments,*

$$\frac{W_n^\mu}{n} \rightarrow Z_\mu := m_\mu(\xi), \quad (40)$$

The distribution of Z_μ is given by

$$Z_\mu \stackrel{d}{=} B_{\mu_0} \mu_0 Z^0 + (1 - B_{\mu_0})(1 - \mu_0) Z^1 + 1, \quad (41)$$

where B_{μ_0}, Z^0, Z^1 are independent and B_{μ_0} has the Bernoulli distribution $B(\mu_0)$. The distributions of Z^0 and Z^1 are the unique solutions of the system (42). The mean $\kappa_\mu := \mathbb{E}[Z_\mu]$ is given in (43).

Remark: As opposed to the uniform model studied in Section 2.1 in which $p_{00} = p_{10} = 1/2$, by Theorem 2.3, the first order term of $Y_n^i(\lfloor tn + 1 \rfloor)$, $t \in [0, 1]$, depends on t in the general model. This explains the randomness persisting on the right-hand side of (40) and accounts for the standard deviation of $Y_n^i(\lfloor tn + 1 \rfloor)$ being of linear order. To be more precise, it is not hard to show that, for any $\varepsilon > 0$ and $t \notin \mathcal{C}_\infty^i$, we have, in probability,

$$\frac{Y_n^i(\lfloor tn + 1 \rfloor) - \mathbb{E}[Y_n^i(\lfloor tn + 1 \rfloor)]}{n^{1/2+\varepsilon}} \rightarrow 0.$$

Thus, the entropy in the choice of the rank in the analysis of grand averages is the only source of randomness of linear order.

Proof of Theorem 3.1. The convergence in probability immediately follows from (24) and the theorem of dominated convergence: for any initial distribution μ and $\varepsilon > 0$, as $n \rightarrow \infty$,

$$\mathbb{P}(n^{-1}|Y_n^\mu(\lfloor \xi n \rfloor + 1) - m_\mu(\xi)| > \varepsilon) = \int_0^1 \mathbb{P}(n^{-1}|Y_n^\mu(\lfloor tn \rfloor + 1) - m_\mu(t)| > \varepsilon) dt \rightarrow 0,$$

since $\mathbb{P}(n^{-1}|Y_n^\mu(\lfloor tn \rfloor + 1) - m_\mu(t)| > \varepsilon) \rightarrow 0$ for almost all $t \in [0, 1]$. The convergence of all moments follows from Theorem 4.5 and the theorem of dominated convergence.

(41) is a direct consequence of (22). $Z^0 := m_0(U)$ and $Z^1 := m_1(U)$ satisfy the limit system

$$Z^i \stackrel{d}{=} B_{p_{i0}} p_{i0} Z^0 + (1 - B_{p_{i0}})(1 - p_{i0}) Z^1 + 1, \quad i = 0, 1, \quad (42)$$

where Z^0, Z^1 and $B_{p_{i0}}$ are independent and $B_{p_{i0}}$ has the Bernoulli $B(p_{i0})$ distribution for $i = 0, 1$. From this system we obtain for the expectations $\kappa_i := \mathbb{E}[Z^i]$ for $i = 0, 1$ that

$$\begin{aligned} \kappa_0 &= \frac{1 + p_{01}^2 - p_{11}^2}{2(p_{00} + p_{11})(1 + p_{00}p_{11}) - 2(p_{00} + p_{11})^2}, \\ \kappa_1 &= \frac{1 + p_{10}^2 - p_{00}^2}{2(p_{00} + p_{11})(1 + p_{00}p_{11}) - 2(p_{00} + p_{11})^2}. \end{aligned}$$

Note that $\kappa_i = 1 + \sum_{k \geq 1} \sum_{1 \leq \ell \leq 2^k} (s_{k,\ell}^i - s_{k,\ell-1}^i)^2$ in the notation of Section 2.2.1. Using (41), we can compute

$$\kappa_\mu = \mathbb{E}[Z_\mu] = \mu_0^2 \kappa_0 + (1 - \mu_0)^2 \kappa_1 + 1. \quad (43)$$

Iterating the system (41) shows that $\mathcal{L}(Z_0)$ and $\mathcal{L}(Z_1)$ both satisfy one-dimensional fixed-point equations. Let $a_1 = p_{00}$ and $b_1 = 1$. Moreover, for $k \geq 2$, let

$$a_k = p_{01} p_{10} p_{11}^{k-2}, \quad b_k = 1 + p_{01} \frac{1 - p_{11}^{k-1}}{1 - p_{11}}.$$

Now, let (A, B) be distributed on $\{(a_k, b_k) : k \geq 1\}$ with $\mathbb{P}((A, B) = (a_k, b_k)) = a_k$. Then $\mathcal{L}(Z_0)$ is characterized by $Z_0 \stackrel{d}{=} AZ_0 + B$. An analogous identity holds for $\mathcal{L}(Z_1)$ upon exchanging the roles of p_{01} and p_{10} (and the other quantities accordingly). It is well-known that fixed-point equations of this type have unique solutions (in distribution) under very mild conditions, compare, e.g. Theorem 1.5 in [20]. This observation concludes the proof. \square

Remark: In the *anti-symmetric* case, that is, $p := p_{00} = p_{11}$, a symmetry argument shows that $\mathcal{L}(Z_0) = \mathcal{L}(Z_1)$, and that this distribution is characterized by the fixed-point equation

$$Z_0 \stackrel{d}{=} (B_p p + (1 - B_p)(1 - p)) Z_0 + 1,$$

with conditions as in (42). This is the same fixed-point equation as in the case $p := p_{01} = p_{11}$. Thus, we know that $Z_0 = m(\xi)$ where m is the linear function given in (26). (Note that this is consistent with Figures 2(b) and 2(d). In both figures, the (closure of the) images of both red and blue functions are equal to the interval $[1/p_+, 1/p_-]$.) In the general case, the limiting distributions are harder to describe. By classical results going back to Grincevičius [9], it is well-known that, under very mild conditions, perpetuities such as $\mathcal{L}(Z_0)$ and $\mathcal{L}(Z_1)$ are either absolutely continuous, singularly continuous or discrete. It is easy to see that both laws are non-atomic, and we leave a more elaborate discussion of their properties for future work.

4 The worst case rank model

We now discuss the worst case rank model.

4.1 Uniform model and asymmetric Bernoulli model

Corollary 4.1. *For the worst case complexity of Radix Selection in the model and notation of Theorem 2.1 we have, as $n \rightarrow \infty$, in distribution and with respect to all moments*

$$\frac{1}{\sqrt{n}} \left(\sup_{1 \leq \ell \leq n} Y_n(\ell) - \frac{b}{b-1} n \right) \xrightarrow{d} \sup_{t \in [0,1]} G(t).$$

Proof. The convergence in distribution follows from Theorem 2.1 and the continuous mapping theorem. For the convergence of the moments note that the proof of $\mathbb{E}[\|X_n^\vartheta - Q_n^\vartheta\|^3] \rightarrow 0$ in the verification of Theorem 2.1 can easily be extended to show that, for any $p \geq 3$,

$$\mathbb{E}[\|X_n^\vartheta - Q_n^\vartheta\|^p] \rightarrow 0.$$

Hence, the limiting behavior of $\mathbb{E}[\|X_n^\vartheta\|^p]$ and $\mathbb{E}[\|Q_n^\vartheta\|^p]$ coincide for any $p > 0$. The proof of Theorem 2.1 also reveals that $(\|Q_n^\vartheta\|)_{n \geq 0}$ is bounded in L_p which yields the assertion. \square

For the Gaussian process G in Theorem 2.1 we have the following results on the tails of its supremum:

Corollary 4.2. *For the supremum $S = \sup_{t \in [0,1]} G(t)$ of the Gaussian process G in Theorem 2.1 we have for any $t > 0$ that*

$$\mathbb{P}(|S - \mathbb{E}[S]| \geq t) \leq 2 \exp \left(-\frac{(b-1)^2}{2b} t^2 \right).$$

Moreover,

$$\text{Var}(S) \leq \frac{b}{(b-1)^2}.$$

Proof. In the context of centered regular Gaussian processes, it is well known that boundedness of the function $t \mapsto \mathbb{E}[G^2(t)]$ leads to bounds on the variance of the supremum S and on its tails. The result follows directly from, e.g., Theorem 5.8 in [2]. \square

Similarly, for $b = 2$ and the asymmetric Bernoulli model, we obtain the following result:

Corollary 4.3. *In the model and notation of Theorem 2.5 we have, as $n \rightarrow \infty$, that in distribution and with respect to all moments*

$$\frac{1}{\sqrt{n}} \left(\sup_{1 \leq \ell \leq n} \left(Y_n^{\text{asyB}}(\ell) - m \left(\frac{\ell}{n} \right) n \right) \right) \xrightarrow{d} S' := \sup_{t \in [0,1]} G^{\text{asyB}}(t).$$

Further, for any $t > 0$, recalling p_+ in (19), we have

$$\mathbb{P}(|S' - \mathbb{E}[S']| \geq t) \leq 2 \exp\left(-\frac{(1-p_+)^2}{2p_+} t^2\right).$$

Moreover,

$$\text{Var}(S') \leq \frac{p_+}{(1-p_+)^2}.$$

Note that Corollary 4.3 deals with the maximum of the centered process. For applications in computer science the worst rank complexity $\sup_{1 \leq \ell \leq n} Y_n^{\text{asyB}}(\ell)$ itself (as for the symmetric case in Corollary 4.1) is of special importance:

Theorem 4.4. *In the model and notation of Theorem 2.5 we have, as $n \rightarrow \infty$, in distribution and with respect to all moments,*

$$\frac{1}{\sqrt{n}} \left(\sup_{1 \leq \ell \leq n} Y_n^{\text{asyB}}(\ell) - \frac{1}{p_-} n \right) \xrightarrow{d} \mathcal{N}\left(0, \frac{p_+}{p_-^2}\right)$$

with $p_- = p \wedge (1-p)$ and $p_+ = 1 - p_-$.

Proof. Assume w.l.o.g. that $p < 1/2$ (otherwise, one exchanges the roles of 0 and 1 in the alphabet). Let $M_n = \sup_{1 \leq \ell \leq n} Y_n^{\text{asyB}}(\ell)$ and $\Delta_n = (M_n - Y_n(1))/\sqrt{n}$. Note that Theorem 2.5 implies the convergence of $(\tilde{Y}_n(1) - n/p)/\sqrt{n}$ towards the normal limit stated above with convergence of all moments by Corollary 4.3. Thus, the statement of the theorem follows from proving $\mathbb{E}[|\Delta_n|^s] \rightarrow 0$ for any $s > 0$. To this end, note that the recurrence (38) yields

$$\Delta_n \stackrel{d}{=} \sqrt{\frac{I_n}{n}} \Delta_{I_n} + R_n, \quad R_n := \max\left\{0, \frac{M'_{n-I_n} - M_{I_n}}{\sqrt{n}}\right\} - \mathbf{1}_{\{I_n=0\}} Y'_n(1), \quad (44)$$

where $I_n, (Y'_k)_{k \geq 0}$ and $(M_k, \Delta_k)_{k \geq 0}$ are independent, $Y'_k = Y'_k(\ell)_{1 \leq \ell \leq k}$ and Y_k have the same distribution for all $k \geq 0$, and $M'_k = \sup_{1 \leq \ell \leq k} Y'_k(\ell)$. Since $I_n/n \rightarrow p$ almost surely and with convergence of all moments, it follows from a standard contraction argument, that the recurrence (44) leads to $\|\Delta_n\|_s \rightarrow 0$ once we have established $\|R_n\|_s \rightarrow 0$. Note that, by Theorem 4.5 in the next section, for all $s > 0$, we have $\mathbb{E}[M_n^s] = O(n^s)$. Since $\mathbb{P}(I_n = 0)$ is exponentially small in n , this implies $\|\mathbf{1}_{\{I_n=0\}} \tilde{Y}_n(1)\|_s \rightarrow 0$. Next, for $0 < \varepsilon < 1/2 - p$, abbreviating $B_n = [(1-p-\varepsilon)n, (1-p+\varepsilon)n]$, we have

$$\begin{aligned} \mathbb{P}(M'_{n-I_n} - M_{I_n} > 0) &\leq \sup_{k \in B_n} \mathbb{P}(M'_{n-k} - M_k > 0) + \mathbb{P}(I_n \notin B_n) \\ &= \sup_{k \in B_n} \mathbb{P}\left(M'_{n-k} - \frac{n-k}{p} - \left(M_k - \frac{k}{p}\right) > \frac{2k-n}{p}\right) + \mathbb{P}(I_n \notin B_n). \end{aligned}$$

By a standard Chernoff bound, the second summand is exponentially small in n as $n \rightarrow \infty$. The first term in the last display converges to zero faster than any polynomial in n^{-1} by Theorem 4.5. Together with the fact that $\mathbb{E}[M_n^s] = O(n^s)$, this shows $\|R_n\|_s \rightarrow 0$ and concludes the proof. \square

4.2 General Markov source model

As in Theorem 2.3 we denote by $Y_n^\mu(\ell)$ the number of bucket operations of Radix Selection with $b = 2$ selecting a rank $1 \leq \ell \leq n$ among n independent data generated from the Markov source model with initial distribution $\mu = \mu_0 \delta_0 + \mu_1 \delta_1$ where $\mu_0 \in [0, 1]$ and transition matrix $(p_{ij})_{i,j \in \{0,1\}}$ with p_+ as in (19). Let

$$M_n^\mu := \max_{1 \leq \ell \leq n} Y_n^\mu(\ell), \quad n \in \mathbb{N}_0, \quad (45)$$

with the convention $M_0^\mu := 0$. For the sake of convenience, we also extend the definition of m_μ in (22) to the entire unit interval by

$$m_\mu(t) = \frac{1}{2}(m_\mu(t-) + m_\mu(t+)), \quad t \in \mathcal{C}_\infty^\mu.$$

Theorem 4.5. *For the worst case rank complexity M_n^μ as in (45) and any $s > 1$, we have*

$$\left\| \frac{M_n^\mu}{n} - \sup_{t \in [0,1]} m_\mu(t) \right\|_s = O(n^{-1/2}), \quad n \rightarrow \infty.$$

In particular, $n^{-1}M_n^\mu \rightarrow \sup_{t \in [0,1]} m_\mu(t)$ almost surely.

The proof of Theorem 4.5 below contains a description of the limiting constant in the theorem:

$$\sup_{t \in [0,1]} m_\mu(t) = (\mu_0 \mathbf{m}_0) \vee (\mu_1 \mathbf{m}_1) + 1 \quad (46)$$

where $(\mathbf{m}_0, \mathbf{m}_1)$ is the unique solution to the system of equations stated in (50). This system can be solved explicitly by a case analysis yielding for $g_1 := g_1(\mathbf{m}_0, \mathbf{m}_1) := p_{00}\mathbf{m}_0 - p_{01}\mathbf{m}_1$ and $g_2 := g_2(\mathbf{m}_0, \mathbf{m}_1) := p_{11}\mathbf{m}_1 - p_{10}\mathbf{m}_0$,

$$(\mathbf{m}_0, \mathbf{m}_1) = \begin{cases} (p_{01}^{-1}, p_{10}^{-1}), & \text{if } g_1 \wedge g_2 \geq 0, \\ (p_{01}^{-1}, p_{10}/p_{01} + 1), & \text{if } g_1 \wedge (-g_2) \geq 0, \\ (p_{01}/p_{10} + 1, p_{10}^{-1}), & \text{if } (-g_1) \wedge g_2 \geq 0, \\ (1 - p_{01}p_{10})^{-1}(p_{01} + 1, p_{10} + 1), & \text{if } (-g_1) \wedge (-g_2) \geq 0. \end{cases}$$

Note that the values of g_1 and g_2 differ in each of the cases since they depend on $(\mathbf{m}_0, \mathbf{m}_1)$. We have, e.g., $g_1 = p_{00}p_{01}^{-1} - p_{01}p_{10}^{-1}$ and $g_2 = p_{11}p_{10}^{-1} - p_{10}p_{01}^{-1}$ in the first case and therefore,

$$g_1 \wedge g_2 \geq 0 \iff (p_{01} - p_{10}^2) \wedge (p_{10} - p_{01}^2) \geq p_{01}p_{10}.$$

To prepare the proof of Theorem 4.5, first note that the system (27) discussed in Section 2.2.2 implies a similar system of distributional recursions for $M_n^i = \sup\{Y_n^i(\ell) : 1 \leq \ell \leq n\}$, $n \geq 2$:

$$\begin{aligned} M_n^0 &\stackrel{d}{=} M_{I_n^0}^0 \vee M_{n-I_n^0}^1 + n, \\ M_n^1 &\stackrel{d}{=} M_{I_n^1}^0 \vee M_{n-I_n^1}^1 + n, \end{aligned} \quad (47)$$

with $(M_k^0)_{k \geq 0}$, $(M_k^1)_{k \geq 0}$ and (I_n^0, I_n^1) independent and $\mathcal{L}(I_n^i) = B(n, p_{i0})$ for $i = 0, 1$. Consider the rescaling

$$V_n^i := \begin{cases} 0, & \text{if } n = 0, \\ \frac{M_n^i}{n}, & \text{if } n \geq 1, \end{cases}$$

and note that (47) implies

$$\begin{aligned} V_n^0 &\stackrel{d}{=} \left(\frac{I_n^0}{n} V_{I_n^0}^0 \right) \vee \left(\frac{n - I_n^0}{n} V_{n-I_n^0}^1 \right) + 1, \\ V_n^1 &\stackrel{d}{=} \left(\frac{I_n^1}{n} V_{I_n^1}^0 \right) \vee \left(\frac{n - I_n^1}{n} V_{n-I_n^1}^1 \right) + 1, \end{aligned}$$

with conditions as in (47). The law of large numbers suggests that limits V^0 and V^1 should satisfy

$$\begin{aligned} V^0 &\stackrel{d}{=} (p_{00}V^0) \vee (p_{01}V^1) + 1, \\ V^1 &\stackrel{d}{=} (p_{10}V^0) \vee (p_{11}V^1) + 1, \end{aligned} \quad (48)$$

with V^0, V^1 independent. In fact, there is a deterministic solution to (48): $(V^0, V^1) = (a, b)$ solves (48) for $a, b \in \mathbb{R}$ if and only if (a, b) is a fixed point to the map

$$T_2 : \mathbb{R}^2 \rightarrow \mathbb{R}^2, \quad (x, y) \mapsto ((p_{00}x) \vee (p_{01}y) + 1, (p_{10}x) \vee (p_{11}y) + 1). \quad (49)$$

Thus, the existence of a (deterministic) solution may be deduced from an analysis of T_2 . More precisely, it is sufficient to show that T_2 is a contraction with respect to a complete metric on \mathbb{R}^2 . This holds for the metric induced by the maximum norm $\|\cdot\|_\infty$ and is deduced from the next lemma.

Lemma 4.6. *For all $p \geq 1$ and $a, b, c, d \in \mathbb{R}$,*

$$|a \vee b - c \vee d|^p \leq |a - c|^p + |b - d|^p.$$

Proof. By symmetry, one may assume without loss of generality that $a \geq b$. The assertion follows immediately when considering the cases $c \geq d$, $(c < d) \wedge (a \geq d)$ and $(c < d) \wedge (a < d)$ separately. \square

Corollary 4.7. *T_2 is a contraction on $(\mathbb{R}^2, \|\cdot\|_\infty)$. In particular, T_2 has a unique fixed point.*

Proof. Lemma 4.6 implies for any $a, b, c, d \in \mathbb{R}$,

$$\|T_2(a, b) - T_2(c, d)\|_\infty^2 \leq ((p_{00}^2 + p_{01}^2) \vee (p_{10}^2 + p_{11}^2)) \|(a - c, b - d)\|_\infty^2.$$

Hence, T_2 is a contraction since $p_+ < 1$ and $p_{00} + p_{01} = 1 = p_{10} + p_{11}$. Moreover, T_2 has a unique fixed point by Banach's fixed point theorem and the completeness of $(\mathbb{R}^2, \|\cdot\|_\infty)$. \square

Proof of Theorem 4.5. Let $(\mathbf{m}_0, \mathbf{m}_1) \in \mathbb{R}^2$ be the unique solution to the system

$$\begin{aligned} \mathbf{m}_0 &= (p_{00}\mathbf{m}_0) \vee (p_{01}\mathbf{m}_1) + 1, \\ \mathbf{m}_1 &= (p_{10}\mathbf{m}_0) \vee (p_{11}\mathbf{m}_1) + 1, \end{aligned} \quad (50)$$

which exists by the previous corollary. Note that the properties (iii) and (iv) of m_0 and m_1 listed on page 10 imply that the unique solution to (50) is given by

$$(\mathbf{m}_0, \mathbf{m}_1) = \left(\sup_{t \in [0,1]} m_0(t), \sup_{t \in [0,1]} m_1(t) \right).$$

Equations (48) and (50) yield for any $p > 1$ and $i = 0, 1$

$$\begin{aligned} \|V_n^i - \mathbf{m}_i\|_p &\leq \left\| \left(\frac{I_n^i}{n} V_{I_n^i}^0 \right) \vee \left(\frac{n - I_n^i}{n} V_{n - I_n^i}^1 \right) - \left(\frac{I_n^i}{n} \mathbf{m}_0 \right) \vee \left(\frac{n - I_n^i}{n} \mathbf{m}_1 \right) \right\|_p \\ &\quad + \left\| \left(\frac{I_n^i}{n} \mathbf{m}_0 \right) \vee \left(\frac{n - I_n^i}{n} \mathbf{m}_1 \right) - (p_{i0}\mathbf{m}_0) \vee (p_{i1}\mathbf{m}_1) \right\|_p \end{aligned}$$

Abbreviate $d_i(n) = \|V_n^i - \mathbf{m}_i\|_p$. Using Lemma 4.6 and the fact that the normal approximation for the binomial distribution $B(n, p)$ is valid with respect to arbitrary moments, it follows that, for some universal constant $C_1 > 0$ and all $n \geq 1$,

$$d_i(n) \leq \left\| \left(\frac{I_n^i}{n} V_{I_n^i}^0 \right) \vee \left(\frac{n - I_n^i}{n} V_{n - I_n^i}^1 \right) - \left(\frac{I_n^i}{n} \mathbf{m}_0 \right) \vee \left(\frac{n - I_n^i}{n} \mathbf{m}_1 \right) \right\|_p + C_1 n^{-1/2}.$$

Again using Lemma 4.6, we deduce

$$d_i(n) \leq \mathbb{E} \left[\left(\frac{I_n^i}{n} \right)^p \left| V_{I_n^i}^0 - \mathbf{m}_0 \right|^p + \left(\frac{n - I_n^i}{n} \right)^p \left| V_{n - I_n^i}^1 - \mathbf{m}_1 \right|^p \right]^{1/p} + C_1 n^{-1/2}.$$

Thus, with $d(n) = d_1(n) \vee d_2(n)$, it follows

$$d_i(n) \leq \mathbb{E} \left[\left(\frac{I_n^i}{n} \right)^p (d(I_n^i))^p + \left(\frac{n - I_n^i}{n} \right)^p (d(n - I_n^i))^p \right]^{1/p} + C_1 n^{-1/2}. \quad (51)$$

Note that, for any $a > 1$, there is a constant $c_a \in (0, 1)$ such that, for all integers $n \geq 2$, we have $\mathbb{E}[(I_n^i/n)^a + (1 - I_n^i/n)^a] \leq c_a$. This is due to the fact that $\mathbb{E}[(I_n^i/n)^a + (1 - I_n^i/n)^a] < 1$ for all $n \geq 2$ and $\mathbb{E}[(I_n^i/n)^a + (1 - I_n^i/n)^a] \rightarrow p_{i0}^a + (1 - p_{i0})^a < 1$.

Since $\mathbb{P}(I_n^i \in \{0, n\})$ is exponentially small in n , a simple induction over n shows that $d(n), n \geq 1$ is a bounded sequence. Again by induction over n , we now show that, for some large $C > 0$ and all $n \geq 1$, we have $d(n) \leq Cn^{-1/2}$. To this end, assume that the statement was correct for all $m < n$. Then, since $d(n)$ is bounded and $\mathbb{P}(I_n^i \in \{0, n\})$ is exponentially small, there exists a universal constant $C_2 > 0$, such that, (51) and the induction hypothesis yield

$$d_i(n) \leq Cn^{-1/2} \mathbb{E} \left[\left(\frac{I_n^i}{n} \right)^{p/2} + \left(\frac{n - I_n^i}{n} \right)^{p/2} \right]^{1/p} + C_2 n^{-1/2}.$$

The right hand side is bounded by $Cn^{-1/2}$ for $C \geq C_2/(1 - c_{p/2})$ proving the assertion for M_n^0 and M_n^1 . The claim for a general initial distribution $\mu = \mu_0\delta_0 + \mu_1\delta_1$ follows from similar arguments relying on (46) and the recurrence

$$M_n^\mu \stackrel{d}{=} M_{K_n}^0 \vee M_{n-K_n}^1 + n$$

with $(M_k^0)_{k \geq 0}$, $(M_k^1)_{k \geq 0}$ and K_n independent and $\mathcal{L}(K_n) = \mathcal{B}(n, \mu_0)$. □

References

- [1] Billingsley, P. (1999) Convergence of probability measures. Second edition. Wiley Series in Probability and Statistics: Probability and Statistics. A Wiley-Interscience Publication. *John Wiley & Sons, Inc., New York*.
- [2] Boucheron, S., Lugosi, G. and Massart, P. (2013), Concentration Inequalities: A Nonasymptotic Theory of Independence. *Oxford University Press*.
- [3] Cesaratto, E. and Vallée, B. (2015) Gaussian Distribution of Trie Depth for Strongly Tame Sources. *Combinatorics, Probability and Computing* **24**, 54–103.
- [4] Clément, J., Flajolet, P. and Vallée, B. (2001) Dynamical sources in information theory: a general analysis of trie structures. Average-case analysis of algorithms (Princeton, NJ, 1998). *Algorithmica* **29**, 307–369.
- [5] Devroye, L. (1986) Lecture notes on bucket algorithms. Progress in Computer Science, 6. *Birkhäuser Boston, Inc., Boston, MA*.
- [6] Devroye, L. (1992) A study of trie-like structures under the density model. *Ann. Appl. Probab.* **2**, 402–434.
- [7] Fill, J.A., Janson, S. (2000) A characterization of the set of fixed points of the Quicksort transformation. *Electron. Comm. Probab.* **5**, 77–84 (electronic).
- [8] Flajolet, P. and Sedgewick, R. (2009) Analytic combinatorics. *Cambridge University Press, Cambridge*.

- [9] Grincevičius, A.K. (1974) The continuity of the distribution of a certain sum of dependent variables that is connected with independent walks on lines *Theor. Probability Appl.* **19**, 163–168.
- [10] Hun, K. and Vallée, B. (2014) Typical depth of a digital search tree built on a general source. *ANALCO14—Meeting on Analytic Algorithmics and Combinatorics*, 1–15.
- [11] Jacquet, P. and Régnier, M. (1988) Normal limit distribution for the size and the external path length of tries. INRIA Research Report 827.
- [12] Janson, S. (2012) Renewal theory in the analysis of tries and strings. *Theoret. Comput. Sci.* **416**, 33–54.
- [13] Kirschenhofer, P., Prodinger, H. and Szpankowski, W. (1989) On the variance of the external path length in a symmetric digital trie. *Combinatorics and complexity (Chicago, IL, 1987). Discrete Appl. Math.* **25**, 129–143.
- [14] Knuth, D.E. (1998) The art of computer programming. Vol. 3. Sorting and searching. Second edition. *Addison-Wesley, Reading, MA*.
- [15] Leckey, K., Neininger, R. and Sulzbach, H. (2014) Analysis of radix selection on Markov sources. *Proceedings of the 25th International Conference on Probabilistic, Combinatorial and Asymptotic Methods for the Analysis of Algorithms*. (Eds. M. Bousquet-Mélou, M. Soria). *DMTCS-HAL Proceedings series*, 253–264.
- [16] Leckey, K., Neininger, R. and Szpankowski, W. (2013) A Limit Theorem for Radix Sort and Tries with Markovian Input. Submitted for publication. See arXiv:1505.07321.
- [17] Mahmoud, H.M., Flajolet, P., Jacquet, P. and Régnier, M. (2000) Analytic variations on bucket selection and sorting. *Acta Inform.* **36**, 735–760.
- [18] Sulzbach, H., Neininger, R. and Drmota, M. (2014) A Gaussian limit process for optimal FIND algorithms. *Elect. J. Probab.* **19**, no. 3, 1–28.
- [19] Szpankowski, W. (2001) Average case analysis of algorithms on sequences. With a foreword by Philippe Flajolet. Wiley-Interscience Series in Discrete Mathematics and Optimization. *Wiley-Interscience, New York*.
- [20] Vervaat, W. (1979) On a stochastic difference equation and a representation of nonnegative infinitely divisible random variables. *Adv. in Appl. Probab.* **4**, 750–783.